HIGHER EDUCATION:
Handbook of Theory and Research

Volume XIII

Associate Editors

HIGHER EDUCATION:

Handbook of Theory and Research

Volume XIII

Edited by

John C. Smart
University of Memphis

Published under the sponsorship of
The Association for Institutional Research (AIR)
and
The Association for the Study of Higher Education (ASHE)

AGATHON PRESS
New York

5648 Riverdale Avenue
Bronx, NY 10471-2106

ISBN: 0-87586-121-0 (paper)
ISBN: 0-87586-122-9 (cloth)

Library of Congress Catalog Card Number: 86-642109

Contents

Contributors

MELISSA S. ANDERSON is Associate Professor of Higher Education at the University of Minnesota. Her recent research has focused on faculty demography, graduate education, scientific misconduct, ethical issues in research, and academic-industry relations.

ALAN E. FECHTER served as President of the Commission on Professionals in Science and Technology (CPST) until his untimely death in April 1997. CPST is a non-profit research group that is a participating organization of the American Association for the Advancement of Science (AAAS). Previously, he was Executive Director of the Office of Scientific and Engineering Personnel of the National Research Council (NRC) (the principal operating agency of the National Academy of Sciences and the National Academy of Engineering). In this role, he oversaw the conduct of the Survey of Earned Doctorates and Survey of Doctorate Recipients. Before joining the NRC, he was Director of Science Resources Studies at the National Science Foundation (NSF), one of the primary funding agencies for data collection on doctorates. Earlier in his career, he worked at the Urban Institute. He was an economist by training.

KENNETH A. FELDMAN is professor of sociology at the State University of New York at Stony Brook. He is a consulting editor for *Research in Higher Education* and has published many articles in that journal; his work has also appeared in a wide array of other professional journals. His book with Theodore M. Newcomb, *The Impact of College on Students,* has been republished as part of a series on the foundations of higher education. With Michael Paulsen, he recently co-authored an ASHE-ERIC Higher Education Report entitled *Taking Teaching Seriously: Meeting the Challenge of Instructional Improvement* and is currently working with him on the second edition of their co-edited ASHE reader *Teaching and Learning in the College Classroom.* A member of the editorial board of the *Journal of Higher Education* since 1994, he is a recipient of the 1994 Wilbert J. McKeachie Career Achievement award of the Special Interest Group in Faculty Evaluation and Development (SIGFED) of the American Educational Research Association, the 1995 American Educational Research Association Distinguished Research Award for the Postsecondary Education Division, and the 1996 Research Achievement Award from the Association for the Study of Higher Education.

CATHERINE D. GADDY has served as Executive Director of CPST since 1995. In this role, she serves CPST's academic, corporate, and professional society members in meeting their needs for data on education and employment of scientists and engineers, and in facilitating discussion of related policy issues; she also performs grant-supported research for AAAS, Ford Foundation, NRC, NSF, and the Alfred P. Sloan Foundation. Previously, she worked for two years for the American Psychological Association, including conducting outcomes assessment research on graduate education programs and graduates. She has served as an adjunct faculty member at the undergraduate and graduate levels, most recently teaching statistics and research methodology in a graduate program. Before joining APA, she worked for 13 years for an engineering consulting company in applied human performance research, development and consulting, primarily for clients in the electric power industry and U.S. Navy. She received her doctorate in psychology from the Catholic University of America.

JAMES C. HEARN is a professor in the Institute of Higher Education of the University of Georgia. Both his research and his teaching focus on organization and policy in higher education. Among his current research projects are an empirical investigation of the changing demography of the professoriate and an analysis of the political history of the federal student loan programs.

CARL J HUBERTY is Professor of Educational Psychology, Measurement, and Statistics at the University of Georgia in Athens, where he has been a faculty member since 1969. He received his doctorate in educational statistics from the University of Iowa. He is the author of *Applied Discriminant Analysis* (Wiley, 1994) and has contributed chapters and articles to many books and journals. He has also conducted workshops and given presentations both in the United States and abroad.

JEROEN HUISMAN is research associate at the Center for Higher Education Policy Studies, University of Twente, the Netherlands. His 1995 dissertation focused on a theoretical analysis of the concepts of differentiation and diversity in higher education. The empirical part of his study related to program differentiation and diversity in the Dutch university sector, 1974 - 1993. His present research interests include curriculum design and development, institutional diversity, and comparative research on the effects of governmental steering strategies.

LARRY L. LESLIE is Professor of Higher Education and Senior Fellow, The Udall Center for Public Policy, University of Arizona. His current research areas include how declining state revenues impact the internal operations of public research universities. He has written in many areas of higher education finance including tuition policies, student aid, tax credits, costs, and special problems of historically black colleges.

LAUREEN L. LOWMAN is a graduate research assistant at the University of Georgia in Athens. She is presently working on her doctorate in the area of research, evaluation, measurement, and statistics, with particular focus on the application of multivariate statistics.

MICHAEL MUMPER is an associate professor of political science at Ohio University. His articles have appeared in the *Review of Higher Education, The Journal of Higher Education,* and *The Journal of Education Finance.* He is the author of *Removing College Price Barriers: What Government Has Done and Why It Hasn't Worked* (State University of New York Press).

RONALD L. OAXACA is Professor of Economics and Senior Research Associate, Economics Science Laboratory, University of Arizona. His current research is on women and minorities in science and engineering, the relationship between faculty research and undergraduate education, labor market discrimination, laboratory tests of job search models, and laboratory experimental evaluation of econometric estimators of structural equations. He has written in the areas of labor market discrimination, union effects on wages, unemployment insurance, minimum wages, and job search.

C. ROBERT PACE is Emeritus Professor of Higher Education at UCLA. Over a span of 60 years he has engaged in studies of college graduates, college environments, college students, and other aspects of higher education. He has lived in Arcata, California since 1993, where he continues his research activities and serves as institutional research advisor at Humboldt State University.

GARY RHOADES is at the University of Arizona's Center for the Study of Higher Education. His research focuses on the restructuring of professional labor in higher education. He has completed a book, *Managed Professionals: Unionized Faculty and Restructuring Academic Labor,* currently under review, and is now editing another book, *Managerial Professionals: The Rise and Restratification of Professionals in Higher Education.* His work on administrative costs has appeared in *The Journal of Higher Education, Thought and Action*, and *New Directions for Student Services.*

BECKY ROPERS-HUILMAN is an assistant professor in the College of Education at Louisiana State University. She received her doctorate in Higher Education Administration from the University of Wisconsin. Her research focuses on feminist teaching in higher education, implications of multiple theoretical approaches for methodological choices, and curricular issues as they intersect with opportunities for equitable and diverse practices. Her book *Feminist Teaching in Theory and Practice: Situating Power and Knowledge in Post-structural Classrooms* is due to be published by Teachers College Press in early 1998.

Recollections and Reflections

C. Robert Pace

[Editor's Note*: The following essay by C. Robert Pace is the first of a new initiative in the* Handbook. *Several distinguished scholars from other disciplines whose cumulative contributions are seminal to the development of higher education research literature have been invited to write autobiographical essays or memoirs recounting their professional careers and their reflections on the field of higher education. Higher education as a field of study was essentially founded by scholars from other disciplines, and those of us actively engaged in the field today owe them a great debt. We have much to learn from their insights over the past four or five decades.*

The authors invited to prepare these essays have been given a free hand in shaping their contributions. They have been asked to focus on two broad themes. First, they were encouraged to share their thoughts on what brought them to higher education as a field and on the major developments and individuals that shaped their careers. Second, they were invited to share their perspectives on the evolution of the field and to offer advice to contemporary scholars on needed research. In essence, they were invited to recount both their "personal" reflections on their careers and their "professional" reflections on the field.

This initial contribution by C. Robert Pace will be followed in subsequent volumes with reminiscences by Wilbert J. McKeachie and Burton R. Clark. We hope you find their essays enjoyable and instructive—JCS.]

In a letter dated January 6, 1996, John Smart invited me to write an autobiographical essay/memoir for inclusion in the *Handbook*. He wished to begin a series of such essays by senior scholars whose activities have shaped the field of higher education research. I was, of course, honored by this invitation and said yes. Then from time to time over the next several months I wondered what to say, and how to say it. Looking back, I saw that my jobs had been the main influence on my thoughts about higher education—spanning a 60-year period beginning at Minnesota in the 1930s and ending at UCLA in the 1990s.

What follows is a personal recollection, because when I moved from Los Angeles to Arcata in 1993 I left nearly all of my professional books to

J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*

the Graduate School of Education at UCLA and many cartons of correspondence, reports, and records of my professional activities for the UCLA archives.

INTRODUCTION TO MEASUREMENT: THE MINNESOTA YEARS

In the fall of 1933 I enrolled as a graduate student in psychology at the University of Minnesota. All of us were introduced to Pavlov's dogs, Thorndike's cats, and Tolman's rats. Many of us also took Donald Paterson's course in individual differences. In the winter quarter one of the courses was Group Mental Tests, taught by Alvin C. Eurich. There were about 60 students in the class. We all took a series of tests—the Army Alpha, Army Beta, American Council Psychological Exam, a mental test developed at Ohio State, a Reading Comprehension exam, and several others. In each case we made a distribution of the scores for the class, computed the mean, standard deviation, percentile ranks, etc., and read the test manual and published norms. This was my first real encounter with measurement.

At the beginning of the spring quarter I went to see Dr. Eurich to ask if he knew of any graduate assistantship that might be available. Dr. Eurich was the assistant director of the University's Committee on Educational Research where several graduate students were employed. A position had just opened. It was to supervise 20 or so clerks who hand-scored objective achievement tests, converted the raw scores to percentile ranks, made item analyses, and prepared other information about test responses to report to the instructors. Most of the tests were from the General College, a new enterprise in the university, where all the courses were new courses, the instructors were borrowed from other university departments, the classes were large, and students' progress was indicated by their performance on comprehensive objective achievement tests. It was an experimental college in every sense, different from all other parts of the university. So, my job with the Committee on Educational Research put me in touch with the latest developments in educational testing and the latest developments in higher education curricula at the university.

Committee on Educational Research

This was a time of great activity in the history of testing. Many types of objective test items were developed—true-false, multiple choice, reverse multiple choice, matching, sequencing, items to measure application of knowledge and principles, interpretation of data, relationships, and more. On all the tests, the instructors received detailed information about students' responses to every item—its difficulty, discrimination, correlation with other items, etc. Such information presumably contributed to the improvement of examinations as well as enabling the instructors to see exactly what knowledge and skills the students had acquired. The connection between testing and teaching was apparent, because the content of test items should reflect the content of the course being taught. The intellectual and creative task of constructing reliable and valid objective test items is one of the best ways of clarifying the objectives of a course.

After about a year the job of supervising test scoring and item analyses was passed on to another graduate student, and that enabled me to work on various research projects in the Committee on Educational Research. Three events were especially important in my education and subsequent career.

In 1935, September 9 to 14, Minnesota held an invitational conference on research in higher education. More than 50 scholars came to this conference. Seminars were held on such topics as curriculum, instruction, objectives, student personnel and guidance, internal organization, and regionalism in higher education. The current state of knowledge was noted, and new research ideas were discussed. The participants were the leading contributors to this research: for example, Herman Remmers at Purdue, Coleman Griffith at Illinois, W.H.Cowley at Ohio State, Earl McGrath at Buffalo, Truman Kelley at Harvard, Fred Kelly at the U.S. Office of Education, and others from Columbia, Yale, Stanford, Chicago, Michigan, Nebraska, and elsewhere. I had been asked by Dr. Eurich to review the research literature on some of the topics and to attend the meetings. A bibliography of more than 250 publications was distributed. In all, the conference was a significant opportunity for me to see the scope of higher education research and get acquainted with some of the people who were doing it.

In 1936 President Coffman told Dr. Eurich that he would like to have the Committee on Educational Research make a survey of the occupational and economic status of Minnesota graduates, especially during the depression years. Between 1928 and 1936 there were about 14,000 bachelor degree recipients whose addresses were available at the alumni office. A one-page questionnaire mailed to each of them simply asked them to list their occupational and educational experience since they received their first degree, to indicate the dates of any unemployment, to indicate how soon after they got their first degree they got their first position, and, with respect to their first position and to their present position, how closely related it was to their field of specialization at the university. I was responsible for managing the survey, analyzing the data, and preparing a report of the results. More than 6,000 questionnaire returns were classified and hand tabulated—by year of graduation, gender, geographic region, and the different schools and colleges in the university such as engineering, business, education, liberal arts, agriculture, etc. The report of this research, *A Follow-Up Study of Minnesota Graduates from 1928 to 1936,* was published by the university in 1938; and the gist of it and its significance were briefly summarized.

> Out of this picture of the trends in job opportunities and the occupational and financial status of recent college graduates should come a more realistic and sounder conception of the function of higher education. A Bachelor's degree is not an insurance policy against the effects of an economic depression. During 1932 and 1933 jobs for college graduates were harder to find, they were less likely to be in line with the student's preparation, they were at a lower occupational level, and they paid less money. If the value

of a college education is conceived in terms of immediate job getting and money making, then education was of limited value to the depression graduates.

Today, 1996, when many people judge the value of higher education by jobs and money, there is again a need for a more realistic and sounder conception of the function of higher education.

The third activity having long-range influence on my development was my Ph.D. dissertation. The topic was the measurement of attitudes. Psychologists typically inferred attitudes from stated opinions. I developed and validated a test to infer attitudes from stated behavior and further studied the relationship between attitudes and information.

My interest in attitude surveys and opinion polls has continued to the present day. For example, in studies of college students and college graduates over many years I have usually included measures of attitudes, and been concerned with how well attitudes predict action. I joined the American Association for Public Opinion Research in 1948 (it was founded in 1947), and have been a regular reader of *Public Opinion Quarterly.* How one thinks about questions and responses in higher education opinion surveys can benefit from acquaintance with the public opinion survey literature.

The General College

The General College at the University of Minnesota was established in 1932 to experiment with a new type of educational program, appropriate for students who wanted to come to the university but who were not eligible for admission to the existing colleges and schools because of low grades and aptitude scores, and the college would be for two years rather than four years. The first curriculum was set up on best guesses about the needs of the students. There were ten areas of study—human biology, physical sciences, euthenics, speech and writing, literature and the arts, psychology, current affairs, social problems, economics, and history and government. The course content was not the usual introduction to an academic discipline; the content was selected for its presumed relevance to the students' needs and interests. In 1935, with a grant from the General Education Board of the Rockefeller Foundation, the college established a student personnel service and research staff, and John G. Darley directed a comprehensive survey of General College students—their skills, interests, attitudes, problems, activities, and needs. This reflected the belief that knowing what to teach and how to teach requires knowledge about the students. Beyond a concern about the immediate interests and needs of current students, more long-range needs must also be considered, needs likely to be important as students become young adults facing problems of out-of-school living. So, a second major study was started—a study of former Minnesota students who were now young adults—which was also supported by the General Education Board of the Rockefeller Foundation. Locally we referred to this study as the adult study. In September of 1937 I became the director of it. With my Ph.D. completed in July, and thanks to the faculty members who had been directing the study—designing the sample, developing the content, etc.—both having left Minnesota for other jobs, the job opportunity presented itself.

The sample for the study consisted of 1,600 names, 800 men and 800 women, 400 from each of the entering classes of 1924, 1925, 1928, and 1929, drawn proportionately from the four largest undergraduate divisions of the university, and selected at random from alphabetical lists within the four colleges. The content of the proposed questionnaire was organized under four broad areas—personal, home and family, vocational, and social and civic. Within each area the instructors were asked to prepare questions about the activities, needs and problems, attitudes or points of view they thought would be valuable in their teaching. The cooperative process of working and thinking to prepare the questionnaire took more than a year. When I became director, about two-thirds of the questionnaire content had been assembled. My job was to finish it, get it printed and distributed, analyze all the replies, distribute the results to all the faculty members, and write a final report of the study.

The final questionnaire, titled "Building the University of Tomorrow," was very attractively printed and illustrated with photographs and line drawings, easy to read, easy to answer in most parts, had altogether about 1,000 items, and filled a 52-page booklet. The questionnaires were mailed December 2, 1937, and three months later, after two follow-up postcards, a two-page follow-up letter, and two more postcards, nearly 70 percent of the adults who received them had filled them out and returned them. I do not know whether such a high rate of return could be obtained today. But I believe that the main factors accounting for the high return then are still valid and influential. In addition to the attractiveness of the questionnaire, the main factors were that it came from an institution held in high esteem and in which they all had a personal experience, and that the content dealt directly with activities, interests, and concerns in their everyday life.

Because getting a 70 percent return to a 52-page questionnaire may be unique in higher education and survey research, more information about it may be useful to present day researchers. At the beginning of the questionnaire booklet there was a message from Malcolm MacLean, director of the General College. Here is some of that message.

> One of the best ways to evaluate education and to plan the education of the future is to discover what former students are doing now and what their experiences have been. That is why we are sending you this questionnaire. We hope you will find it interesting and challenging, and that you will derive as much benefit from the chance to think about your activities, problems, and points of view as we will derive from examining the results which you and 1600 other young men and women may send to us.... During the winter and spring, we hope we may have an opportunity to talk personally with many of you who are living in Minnesota, provided, of course, that you are willing. If you are willing to help us further through an interview about some of your activities and problems, will you write your name and address on the following lines?

More than half of the questionnaires were signed, including more than 100 by people who were not even living in Minnesota. From about 260 Twin City residents we interviewed 172. Almost all of them had shown the questionnaire to their friends and had talked about it; said they found it interesting and stimulating; said it took a long time to fill out but was a pleasure; and liked the attractive way it was put together. Nearly 100 of them asked if they could have copies of it to keep.

My advice to questionnaire makers is this: If you really put the time and effort and thought that is needed to construct a good instrument, one that is clear and can be answered from experience, that deals with topics of definite relevance and importance, and that is attractively presented, the chances are fairly good that those who get it will answer it.

From the experience of analyzing and reporting the answers to nearly 1,000 questionnaire items I came to some conclusions about research methods which I thought then and still think today were useful. The sample of names for the survey had been drawn from entering students and so, of course, about half of them subsequently graduated and half did not. One could compare graduates and non-graduates, all of whom were initially qualified for admission to college and all of whom began college. A control group like this may be impossible to get today because a much higher percentage of students eventually graduate, and many of them attend more than one college. In the adult study the questionnaire respondents were divided into eight groups and the same complete analysis of responses was made in each group. The eight groups were identified first between men and women, then for each of these two groups the cases were divided by year of entrance to the university (1923-4 and 1928-9), and then each of those four groups was divided between graduates and non-graduates. Within each group we calculated the percent of adults giving a particular response to every one of the nearly 1,000 items.

One of the advantages in this procedure was to see what general consistency there might be in the responses. Was there a significant difference between graduates and non-graduates on all comparisons or only on one or two? At no time did we ever consider percentages that were based on the total number of respondents. The responses of all groups to all items were reported to all General College instructors and staff members. The detail was in one sense necessary and in another sense overwhelming. One of my contributions to the content of the questionnaire was to include several measures for which normative or other comparative data was available so that scores could be obtained. These included a job satisfaction score, a measure of the economic and cultural status of the home, a liberal-conservative attitude scale, and a measure of general adjustment and morale. Ten years later when I again had a chance to do a follow-up study of former college students nearly all the content of the questionnaire I developed consisted of scales and other sets of items that could be scored.

This is not the place to report the results of the adult study. I will say only that except for occupational level, income, and job satisfaction, the differences between graduates and non-graduates in other aspects of life were not very obvious and frequently non-existent. The story of this adult study—why, how, and with what result—is contained in a book I wrote, *They Went to College*, published by the University of Minnesota Press in 1941.

FROM MEASUREMENT TO EVALUATION: THE COMMISSION ON TEACHER EDUCATION

I left Minnesota in the fall of 1940, having been offered a post-doctoral fellowship by the Commission on Teacher Education, paid by the General Education Board of the Rockefeller Foundation, to spend at Teachers College, Columbia. The Commission, located in the offices of the American Council on Education in Washington, D.C., was embarked on a cooperative study with colleges and school systems for the improvement of teacher education. In the participating institutions a lot of attention was paid to evaluation. Maurice Troyer on the Commission staff was the head of the evaluation activities.

Much of what we did at Minnesota was evaluation and properly called evaluation. Ruth Eckert was employed to do an overall evaluation of the general college programs, and during my last year at Minnesota I worked with her on several projects. In all major areas of the curriculum students took achievement tests at the beginning and the end of the year and in some areas they took the achievement test a year after they had taken the course to see how much had been retained. There were also measures of reading comprehension and of effective writing, and of attitudes, interests, values, vocational plans, and more. The evaluation of student learning and development at the General College was systematic, comprehensive, and reliable. My particular experience at Minnesota, however, was mainly to carry out a specific project.

With the Commission on Teacher Education, the focus was on providing service to the people in the colleges and schools who wanted to evaluate what they were doing. The role was that of a consultant. In the beginning, at Teachers College, the fellowship allowed me to do some things on my own, so I could get better acquainted with teacher education and improve my understanding of measurement and research. I attended the Foundations course that all students had to take. With Irving Lorge, who invited me to come to his course on measurement, I had many discussions about research and evaluation. For the Commission I constructed an exam for the Foundations course, made visits to several colleges to discuss their evaluation activities—the normal school in Oneonta New York, Syracuse University, and Buffalo State College—and attended a workshop where representatives from all the cooperating colleges discussed their evaluation problems. At the end of the year in New York I was invited to join the Commission staff in Washington, D.C., where for the next two years Maurice Troyer and I handled the evaluation activities of the Commission.

In my files at home I did not find a list of all the colleges that were involved in the Commission's work. I know that Dr. Troyer and I visited most of them several times, serving as consultants on the evaluation activities they were undertaking. Among the colleges were Ohio State, Wayne State, Nebraska, Chicago, Texas, Milwaukee, Stanford, Furman, Southern Illinois, William and Mary, Michigan State, and Troy Alabama. We organized workshops and conferences, facilitated communication among the colleges, assembled files of evaluation instruments. Among the evaluation techniques were profiles, rating scales, check lists, essays, achievement tests, case studies, guides to classroom observation, diaries, etc. My experience with the Commission enabled me to become acquainted with institutions and individuals all across the country and see the myriad methods and materials being used in evaluation.

For me, the most important lesson I learned was an attitude or philosophy about how education is improved. This point of view was described briefly in the book Maurice Troyer and I wrote as follows:

> The cooperative study of teacher education emphasized implementation rather than research and survey. Its purpose was to work with groups on their problems rather than to organize a program or to present formal recommendations. It was to work through and to develop local leadership, and to stimulate thinking and experimentation on basic problems in teacher education. Because there were obstacles to the free play of local initiative, the staff of the cooperative study became increasingly interested in how changes were brought about—in the strategy of effective planning. They and the Commission believe that, in the long run, greater progress in teacher education would be made by each institution striving to improve its own program than by any national organization trying to lay down standard recommendations. Thus, in the cooperative study, responsibility for the improvement of programs and procedures remained within each institution. All this had a direct bearing on the philosophy and services of the evaluation division. It meant that there were no comparative studies, no tests centrally developed for use by all the cooperating centers, no prescribed sequence of steps for each to follow in evaluating itself. There was, in other words, no evaluation of schools and colleges by an external agency; rather, there was a working with schools on evaluative tasks chosen by them.

If one purpose for evaluating a program is to improve it then how the evaluation is conducted becomes important. The people whose actions are necessary to change a program need to be active participants in the evaluation. It is sometimes easy to reject recommendations from an evaluation made by someone else but not by an evaluation they themselves have made. Evaluation needs to be a cooperative, collaborative undertaking. The process may determine what is done with the product. Evaluation should not only contribute to learning; it should itself be a learning activity. The simple fact that evaluation has human consequences means that how people are treated will influence what they learn from evaluation. What does a student learn from the teacher's evaluation? What does a teacher learn from the evaluation made by a supervisor? Has the evaluation contributed to learning?

The final report of the evaluation activities includes extensive examples of what was done in the colleges in their teacher education programs—evaluation related to selection and admission, orientation programs, general education, professional education, practice teaching, follow-up, and evaluation in-service on the job. This sequence is one way to view an institutional evaluation plan, with concern about the consistency and relationships among the stages. Another integrating focus is a cumulative record kept by the student. This is described as self-evaluation under guidance. In today's language this would be described as a portfolio.

The philosophy and practice of evaluation exemplified in the work of the Commission is very similar to what is being advocated today under the label of assessment. Both are seen as avenues for learning and development, for change and improvement. And how the work is conducted is probably as important for assessment as it was for evaluation. Our own views and the record of what the colleges did is reported in the book Maurice Troyer and I wrote, *Evaluation in Teacher Education,* published by the American Council on Education in 1944.

OPINION POLLS AND ATTITUDE SURVEYS: THE BUREAU OF NAVAL PERSONNEL

Lieutenant Commander Alvin C. Eurich was the officer in charge of the standards and curriculum section in BuPers. He asked me to join the staff. I did, not as an officer, but as a civilian scientist in the civil service. This activity in the Bureau constructed the Navy's general classification test, and other personnel selection instruments such as reading comprehension, arithmetic, mechanical aptitude, etc.; wrote training manuals and achievement tests for Navy specialties such as Gunner's Mate, Electrician's Mate, etc.; and carried out studies predicting performance from test data, evaluated the instruction and utility of various training programs, and kept in touch with similar programs in the Army and Air Force. I worked in the research unit of this Navy program.

After a year or so most of my work consisted of developing, analyzing, and reporting the results of Information Surveys which I initiated. These were opinions and attitudes of enlisted men regarding their training, and other aspects of their naval experience. The first of these Information Surveys asked men for their opinions about the training they received in their specialty. In other surveys there were items about such topics as job satisfaction, attitude toward officers, pride in outfit, Navy fairness, efficiency, importance. The most extensive survey was an evaluation of the educational services program in the Navy which was under the direction of Lieutenant Commander Earl McGrath. This program included off-duty classes, courses available from the Armed Forces Institute, getting information to ships and bases about progress in the war. The questionnaire was given to samples of officers and men in eight advanced bases in the Pacific, personnel on nearly all types of ships mostly from the Pacific, and the ship's company from one training center in the U.S. A series of nine reports was produced—Getting the News of the Day, Interest in Talks and Discussions about the War, Newsmaps, Appraisal of the Navy's Role in Keeping Men Up-to Date on the News, Participation in USAFI, Participation in Off-duty Classes, Post-War Educational Plans, Accreditation Services, and Interest in War Orientation Readings. At the end of the war we made a survey of post-war plans of reserve officers and of enlisted men (stay in the Navy? get a job? go to school?) The results of that survey, combined with a similar one in the Army, were used to estimate the number of veterans who would go to college.

In the Pentagon the Army had a survey research program. The leaders of it were Samuel Stouffer and Paul Lazarsfeld, and one of its research experts was Louis Guttman. I had known Guttman when he was an undergraduate at Minnesota. In the Pentagon he was refining and developing his theory of scales. A scale is a hierarchical universe of content, a set of items that go together because they come from a unidimensional content domain. Unlike classical test theory, which is a theory of responses, Guttman's scale theory is a theory of content. I made many trips to the Pentagon to get better acquainted with all the opinion surveys that were being developed, and to fully understand what Guttman was doing so that I could make Guttman scale analyses of some of the Navy's opinion polls. In most of the Navy data I did not find scales that met Guttman's criteria; but I was intrigued by the idea and shortly after leaving the Navy I developed scales successfully for a study of college graduates and have continued to the present to develop measures that have the basic characteristics of a scale.

The philosophy of participatory evaluation that was so dominant in the Commission on Teacher Education does not fit the Navy or any other organization that has a firm chain of command. In the Navy, if you want a ten percent sample of personnel at some location to answer a questionnaire, you order it through the proper channels, and you get it. However, you have to be very specific about how to do it. For the Information Surveys we issued nearly five pages of detailed instructions about who should do what, and how, whenever we made a request. This experience was a good lesson on the importance of collecting data under comparable conditions, especially when it is collected by different people at different places.

The other lesson from my Navy experience grows out of the fact that our surveys dealt with many aspects of Navy life and one can see even in a very structured and compartmentalized environment how opinions converge and diverge. Attitude toward officers can be quite positive in some respects and negative in others. Favorable opinions about training programs depend on the connection between school and subsequent job demands, and on having had plenty of opportunity to practice skills not just read about them. Also training programs that are generally good also had instructors who were rated as good. Knowledge about the war such as from the educational services programs contributes to morale, to the belief that what the Navy is doing is important, and that the world might be better after the war. Many factors contribute to job satisfaction. So, in my later research I have tended to look at events in a large context rather than specifically, one at a time.

The final experience from my Navy Department days relates to the creation of the Office of Naval Research. My Minnesota classmate and close friend, Jack Darley, was involved with a few others in drawing up the plans for the Navy to contract with universities for conducting research of interest to the Navy. Jack discussed the plans with me from time to time. ONR was established and a review committee to read and recommend proposals was named. I was designated as the official representative on the committee from the Bureau of Naval Personnel. I knew many of the psychologists on the committee—Jack Darley, Rensis Likert, Lowell Kelley, Carol Shartle, Donald Marquis—and others on the committee I knew only from reading but not in person—Margaret Mead, Ruth Benedict, Erich Fromm. At our first meeting, after we asked Ren Likert to leave the meeting, we approved his proposal for studies of leadership, supervision, and productivity with a long range grant that made possible the development of the Survey Research Center at the University of Michigan.

As that Center grew and expanded over the years it became, in my view, one of the two or three most productive and influential centers of social science research in the country.

Developing Scales for Higher Education

After the war, during the six months between my departure from the Navy Department and my arrival as a faculty member at Syracuse University, I had an opportunity to design an alumni survey for American University. Remembering the complexity of the Minnesota study and my recent knowledge about Guttman scaling theory and methods, I decided to see if I could construct scales for the alumni survey. These would measure aspects of adult life that could be related to general education—to humanities and arts and sciences. For each of seven areas of adult experience— political affairs, civic affairs, art, music, literature, science, and religion—I wrote check-lists of participation. For each topic there were ten activities and alumni were asked to check each activity they had engaged in during the past year. The activities in each topic ranged from ones that were commonplace to ones that suggested a greater level of interest, and that the uncommon activities subsumed participation in all the more common ones. In other words the content of each set of activities would form a unidimensional and hierarchical universe of content as in a Guttman scale.

Here is a brief example: if a person contributed a regular sum of money to the church, you can be pretty sure that person is a member of the church, and attends church services; or if a person says he voted in the last primary or local election, you can be pretty sure that person also voted in the last national election, has listened to political speeches, reads about current events in the paper, and talks about politics with friends. In analyzing the responses to each of these ten item activity check-lists, I found that usually seven or more of them did form a scale as defined by Guttman. So, I learned that I could develop scales to measure some important outcomes of higher education. This had a big influence on much of my future research and on my views about measurement. In achievement testing the length of the test contributes to its reliability. Tests of 100 or more items are typical. With a good scale, however, one can get good reliability with fewer than a dozen items. This means that the range of ones inquiry can be greatly extended and still be feasible administratively. It also means that one must carefully define the concept being measured and determine empirically whether it is a scalable universe of content.

TEACHING, RESEARCH, ADMINISTRATION: THE SYRACUSE UNIVERSITY YEARS

When I went to Syracuse in September of 1947 it was my first employment as a university faculty member. The trustees and administration of the university had decided to undertake a comprehensive survey of its programs and services that would be useful in planning for its future development and needs and that this should be a self-survey. Maurice Troyer, head of the Evaluation Services Center at Syracuse, was asked to be director of the self-survey. At the same time, Dr. Troyer was developing a graduate program in higher education. The need for research in the self-survey plus the need for teaching courses in higher education created the job opportunity for me.

The Self-Survey

The self-survey dealt with a broad range of topics: curriculum and instruction, personnel services, administrative organization, plant and facilities, finances, research and scholarly productivity, library, public relations, with each topic studied by a committee. Every committee included faculty members selected by the academic senate, administrators, and a trustee. In going around to many committee meetings, I could help them gather the information they wanted, merge overlapping requests, devise questionnaires as needed, etc. Altogether at least 100 faculty members and administrators participated in this work, and without any released time from normal responsibilities. The Evaluation Center staff administered the Cooperative General Culture Test and the Current Affairs test to samples of sophomores and seniors. Faculty and students indicated how they rated the importance of each of a list of 18 objectives of general education, with students also indicating how much they thought they were getting toward these goals, and faculty indicating how much responsibility their area assumed for helping students attain the objective. And other systematic data gathering surveys and ratings were undertaken. Each committee prepared a report of its inquiry and its recommendations during the summer of 1948. These reports added up to nearly 1,000 pages. At that point I was asked to prepare a condensed report of less than 100 pages which would accurately reflect what each committee did and which could and would be read by all the members of all the committees, and ultimately by all the faculty members and administrators in the university, and would clearly indicate all the important recommendations that were made. In the fall of 1948 representatives of the survey committees met with the University's top administration for three days to discuss the findings and lay plans for translating recommendations into action. At this point the public relations office argued that the reports should be carefully guarded lest the Syracuse newspapers would write articles about all the things that were wrong at the university. Happily, most of the people at the meeting rejected that view. Ultimately, a 75-page self-survey report to faculty was attractively printed and very widely distributed.

What a wonderful way to begin a career as a professor of higher education! In just over a year I became acquainted with all the university's administrators, many of its faculty, some of the trustees, and they with me. I learned how the place operated—not only its curriculum and instruction, but also its financial structure and needs, its administrative organization, graduate and research programs, student personnel services, its plant and facilities, and other features.

The Chancellor's Office

Not long after completion of the self-survey the Chancellor, William P. Tolley, asked if I would be willing to spend part of my time working in his office as a Special Assistant to the Chancellor. I said OK provided he understood that I would be a special assistant, not an administrative assistant or an expediter or a gopher. For the next three years I had a chance to write some speeches for the Chancellor, assemble materials for his annual reports to the Board of Trustees, answer some of his mail, and participate in many discussions about university policies, personnel, and budgets, and at his request I talked with Deans and Directors about their aspirations and their budgets.

Before WW2 Syracuse had been mainly an undergraduate institution with an up-State Methodist clientele. After WW2 the University, under the Chancellor's leadership, rapidly grew into a comprehensive research university, with a diverse student body, an aggressive research enterprise, and a big-time athletic program. I sat in on many discussions where these decisions were made— how to organize a research institute, how to determine and distribute overhead, etc. There had been no institutional research office so I instituted various data sets such as information about class size, teaching loads, faculty ranks, instructional costs. Once, when the Chancellor was out of town for more than a month prior to the annual meeting of the Trustees, I wrote the report from beginning to end, and the first time he saw it was when his secretary handed him a copy at the Board meeting. The opportunity to see how the university really worked, from inside a chancellor's office, what problems and issues were discussed, how decisions were made, and who made them, all added greatly to my knowledge about higher education. I saw that the knowledge and experiences that characterize life in a chancellor's office are very different from the day-to-day knowledge and experiences of faculty members. I have subsequently felt that misunderstandings between administration and faculty are more often owing to differences in experience and responsibility than to differences in values.

The Evaluation Center

The Evaluation Services Center had played a significant role in the conduct of the self-survey. The Center had a director and an associate director, part-time, a full time secretary, and three or four graduate student research assistants, mainly funded by the Chancellor's office. Several research inquiries for the self-study were carried out as doctoral dissertations by research assistants in the Evaluation Center. After I had been at Syracuse two years, Maurice Troyer left to become vice president of a new International Christian University in Japan, so I became director of the Evaluation Center. The research assistants in the Center were always doctoral candidates in the graduate program in higher education. There were four basic courses in the doctoral program—history, administration, curriculum, and evaluation. I always taught the evaluation course, and sometimes the history or curriculum course. Altogether at Syracuse I supervised twelve dissertations in higher education, six of them by research assistants in the Evaluation Center. In addition to the studies of objectives and achievement tests previously noted there was a study of religious beliefs and social values and a study of Syracuse alumni.

The alumni survey questionnaire was a revision and extension of the questionnaire developed at American University. There were Guttman-type activity scales related to Politics, Civic Affairs, Religion, Literature, Music, Art, and Science, with each scale having eleven items. Then there were opinion measures related to Politics, Civic Relations, Government, The World, Philosophy, Literature, Music, Art, and Science, with each measure having six statements. And, there was the same list of eighteen objectives that had also been used in the studies of students and faculty. The survey was mailed to graduates of the class of '47 and to classes at five year intervals back to '27. The response rate was 52 percent.

With respect to the eighteen objectives there was substantial agreement among students, faculty, and alumni as to their importance—the rank order correlations ranging from .81 to .91. There was also high agreement between students and alumni regarding their progress toward the objectives; and between the ratings of importance and progress in both groups. The alumni survey scores on the Guttman-type activity scales and the opinion measures were analyzed in relation to their college major field, and in all instances the pattern of results for adult activities and opinions was congruent with the undergraduate major field. For example, adults who had been Fine Arts majors were also the highest participants in activities related to the arts and in having opinions similar to experts in the arts. This same clear-cut relationship was found when students' scores on the Cooperative General Culture Test were analyzed—high achievement was always related to the students' major field.

With respect to religious beliefs and social values there was no difference by majors among freshmen, but among seniors the highest scores were made by students in the humanistic and service-oriented curricula and the lowest scores by students in the scientific and technically oriented curricula, so that one would conclude that scores do not seem to be incidental products of just any kind of college program. On all the other studies of students and alumni it seemed clear that college makes a difference and the particular kind of college education also makes a difference, and that the pattern of activity and opinion among alumni is remarkably similar to the pattern among students.

The network of associations that emerges from these studies is still a basically correct picture of college influences on student outcomes. Perhaps one of the reasons for this stability is the reliability of the measurements. My personal interest in measurement has always reflected my belief in the importance of reliability and validity. For the Evaluation Center studies, here are some examples. In the alumni survey we had a sample of respondents answer the activity scales and opinion measures six months later. The test-retest correlations ranged from .83 to .89 for the activity scores and, for seven of the nine opinion measures from .60 to .70 with two low ones. For the activity scales 85 percent of the responses were identical on the two occasions, and for the opinion items 75 percent were identical. On the measures of religious beliefs and social values, test re-test correlations were .93 and .87. Validity is apparent from the content of the measures, but also from the fact that results fit into previously known and predicted outcomes.

The Psychology Department

In the fall of 1952 my role at Syracuse changed. I was chosen to be the chairman of a newly formed psychology department. Prior to that time psychologists at Syracuse were located in various parts of the university—some in Liberal Arts, some in Education, one or two in the School of Business, and one in the Maxwell School of Citizenship. Now all the psychologists would belong to a single all-university department. Psychology at Syracuse was not accredited in anything—not in clinical or counseling, not by VA or Mental Health—and there were no regular faculty members in experimental or physiological psychology and no laboratories.

So the job was to help construct a balanced department, become fully accredited, and promote research. Fortunately, this was a period in higher education of growth in magnitude and in the support of research. At Syracuse three people would be retiring within a year or two. We were authorized to hire replacements and to seek two more. I got money from the National Science Foundation for laboratory equipment. In less than two years we had some outstanding young scholars on the faculty; in three years we were fully accredited in everything, we had many research contracts and grants; and within five years we were rated as among the top 20 departments in the country.

I served as department chairman for nine years. We were always well represented at the annual meetings of the American Psychological Association, and the American Educational Research Association. Locally I spent much of my time just talking with every member of the department so as to know fully what research was being carried out and why it was important. We had monthly departmental meetings where procedures as well as policies were openly discussed and voted on. The larger success story of the department was owing to the fact that good people could be hired and money could be obtained for research grants, training programs, and student support, not to any virtues I may have had as an administrator.

When I indicated that I no longer wished to be chairman it was agreed that I would spend my final year working with the administration and the faculty in a national search for a new chairman. This was a very active search, to which everyone in the department devoted a great deal of thought, and which resulted in coming down to three excellent choices that we were ready to send to the Dean with our recommendations in rank order. The Dean, of course, had met all of them when they had come to the campus for interviews. Then the Dean arbitrarily named a person who had not been a candidate or even indicated an interest in the position, and who was known by but not wanted by most members of the faculty. Moreover, the Dean never met with the department to explain his action. As a consequence nearly all the outstanding people who had been hired left Syracuse within the next few years for positions elsewhere and the department's strong reputation rapidly disappeared. I report this event as an illustration of the importance of good faith between faculty and administration, and the consequences of bad faith.

National Connections

During the years I was at Syracuse I had the opportunity to work with several national organizations—the College Entrance Examination Board, the American Council on Education, the Social Science Research Council, the Fund for the Advancement of Education of the Ford Foundation, and the Carnegie Corporation. Frank Bowles, president of the CEEB, invited me to be a member of the Board's research advisory committee. In the past, all of the research activities of interest to CEEB were performed by the Educational Testing Service. I persuaded the Board that, at least from time to time, it should solicit research proposals from people who had no connection with ETS. One year I submitted a proposal that was funded, and Anne Anastasi also received a grant. I also suggested to the Board that it should have a few people on its research advisory committee who were not from the educational testing establishment and I recommended that Paul Lazarsfeld be invited to bring a sociological and societal perspective to the Board's programs.

For several years I was a member of the American Council on Education's Committee on Measurement and Evaluation. This committee recommended that the Council approve the creation of the Educational Testing Service. It was this committee that also sponsored the Cooperative Study of Evaluation in General

Education, and appointed Paul Dressel to direct the study. A major product of that national study was the development of several tests of critical thinking. These tests represented the thinking of many faculty members about what was meant by critical thinking, and what test items would adequately reflect it—in social studies, communication, sciences, and humanities. People today who think that objective achievement tests can only measure facts are surely not aware of the report of the Cooperative Study by Paul Dressel and Lewis Mayhew.

My notes at home do not indicate when the Social Science Research Council's Committee on Personality Development in Youth was formed. At any rate I was a member of the committee for the duration of its existence. Ralph]Tyler was the chairman and the other members included Dana Farnsworth, Chester Harris, Don Marquis, T.R.McConnell, Lloyd Morrisett, Ted Newcomb, Nevitt Sanford, and Robin Williams. A major purpose of the committee was to interest a variety of social scientists in the study of personality development during the college years. The committee provided some financial support to Nevitt Sanford for his book on the American College, and to Chester Harris for a book on the measurement of change. The committee organized a three-day conference at Amherst to which more than 50 researchers were invited to discuss the present state of knowledge and encourage new lines of inquiry. The committee also enabled me to hold a three-day meeting at the Council's offices in New York to discuss personality measurements for outcomes of higher education, with the following invitees: Ed Bordin, Ben Bloom, Jack Darley, Jane Loevinger, Don Marquis, Lloyd Morrisett, T.R.McConnell, Dave Saunders, George Stern, Harold Webster, Dick Christie, and Robin Williams. Many personality measures are not suitable for measuring outcomes of higher education because they presumably measure stable traits. At the meeting we made a list of important objectives of higher education and for each we listed measures that were available for those objectives. The use of relevant personality measures has an important influence on higher education research because it turns attention to the total college environment—to outcomes that do not come from courses exclusively to ones that look beyond courses and curricula to other features of college life. This wider look at the total college environment can enrich our understanding of higher education. Subsequent events demonstrated that it did.

In 1951 the Ford Foundation established the Fund for the Advancement of Education, with Clarence Faust as president and Alvin C. Eurich as vice president. Over the next several years I had many contacts with Al Eurich—discussing policies and programs with him, and participating in some of the Fund's projects. My own views often differed from those of the Fund, but that never interfered with the frequent support I got from the Fund. I tried to interest the Fund in setting up some major research centers to focus on bringing together various behavioral scientists in studying major educational problems, but the Fund was more interested in national demonstration projects such as the improvement of teacher education, the use of television in the schools, early admission and advanced placement in college, promoting equal educational opportunity. Moreover, the possibility of connecting education and the behavioral sciences was ended when the Ford Foundation decided to abolish its Behavioral Sciences Division.

I contributed to several Fund projects. For a number of years money was given to colleges in Arkansas to raise the level of liberal education of prospective school teachers. As the program was ending the Fund offered small planning grants to the colleges to think about how they might maintain some of their activities. I was asked to meet with the presidents, deans, and others to talk about the values that might be gained from these local grants, and then to visit many of the colleges in Arkansas to answer questions, review their plans, and make suggestions. The experience enlarged my knowledge and appreciation of the diversity of higher education in the country.

Another program was the college teaching internship program where several colleges agreed to employ graduate students who were near to completion of their dissertations and give them an introduction to what college teaching was all about, and what it was like to be a faculty member. They visited classes of other professors and discussed their observations, and in most places they also had seminars about higher education, student characteristics, learning, counseling, etc. At the end of the year they were asked to write their observations and thoughts about teaching; and first year instructors who were not interns were also asked to comment on their year's experience. I was asked to read and summarize all these expressions. The task required familiarity with the methodology of content analysis. What impressed me was the difference in attitude between the interns and the regular instructors, with the interns being much more aware and appreciative of student interests, motivations, and personalities. For the colleges the main impact was that college teaching became a normal topic of conversation among the faculty whereas previously faculty members rarely talked about what they did in their classes.

The Fund gave a grant to Sweet Briar college to evaluate the Junior Year in France program which was administered from the college, and which prior to the war had been administered by the University of Delaware. I was asked to prepare a questionnaire to be sent to alumni of this program, the main purpose of which was to increase international understanding. I developed and pre-tested Guttman-type scales on six topics—international activities, language and cultural activities, policies regarding the exchange of people and information, the role of the United Nations, and policies of the U.S., and acceptance of people of other cultures. There was also a short test designed to tap knowledge about other cultures and countries by asking people to think of the names of outstanding contemporary contributors to literature and the arts, science, philosophy, and other fields from the United States, Great Britain, France, and any other countries. So here was another opportunity to indulge my interest in developing new measures of important variables that had not previously been measured.

In the post-war period there were 14 colleges that had sent 10 or more students to the program, and from those colleges we got a control group of similar liberal arts students. The results of the study were clear. Compared with their contemporaries from the same colleges, alumni of the Junior Year in France program were: 1) personally more tolerant in their acceptance of people who differ from themselves; 2) more fully aware of significant intercultural contributions to life in the twentieth century; 3) more frequent and more active participants in internationally-oriented activities both of a political and cultural sort; and 4) more inclined to endorse policies which promote the freer exchange of ideas, goods, and people among countries. These differences could not be attributed to travel by itself, or to language majors, or to gender. My report of this study, *The Junior Year in France*, was published by the Syracuse University Press in 1959.

Beginning a New Line of Research

The combination of scores on the Scholastic Aptitude Test and high school grades usually predicted college success (freshman grades) by a correlation of about .50. Adding other variables to the prediction equation did not seem to increase the correlation by any great amount. As these studies were discussed in the research advisory committee of the College Board, I wondered whether the relationship might be improved by expanding the criterion end of the equation. Perhaps one could think of college success as involving something more than a grade point average. The Board appointed a subcommittee of Quinn McNemar, Anne Anastasi, and me to think about non-intellectual factors related to college success, including factors in the environment as well as in the individual.

One of the new psychology department faculty members at Syracuse was George Stern, who was familiar with Henry Murray's theory of personality needs and environmental press. A press is an aspect of an environment that is compatible with a personality need. A person with a need for order would like an orderly environment, for example. Performance is best when need and press are similar. Stern had constructed a questionnaire to measure 30 of the personality needs described by Murray. So we proposed to the College Board to construct a set of environmental characteristics parallel to the personal characteristics in Stern's questionnaire. With funds from the CEEB we constructed the College Characteristics Index (CCI) and tried it out on samples of students at five very different colleges. Responses to some of the statements of press were clearly different between some of the institutions, and this suggested the possible value in further exploration of the need-press idea. Much more information was needed about the presumed parallelism between needs and press, for example how well can it be demonstrated empirically. And more thought needed to be given to the content of the CCI items.

After the completion of the report to CEEB I got a discretionary grant from John Gardner at the Carnegie Corporation that enabled work to continue on the revision of the CCI and the collection of more data. Meanwhile I became increasingly uneasy about interpreting the CCI as if college environments had the same characteristics as personalities. We revised the CCI twice and by the summer of 1959 we had data from about 50 colleges and universities. By then it was also clear that Stern and I had very different ideas about how the CCI should be analyzed and interpreted. Stern analyzed student responses to the CCI undifferentiated as to the institution they were describing. My interest was in differentiating between institutions so that the college, not the student, was the proper unit of analysis. Also, correlations between pairs of need-press responses revealed that in quite a few cases they were not parallel. In one or two cases the correlation between need and press was negative. Empirically the presumed relation between need and press was not confirmed in some of the comparisons, and therefore it was not appropriate to report all the environment characteristics as if they were parallel to the personal needs characteristics.

From September 1959 to September 1960 I was a Fellow at the Center for Advanced Study in the Behavioral Sciences. During that year I analyzed institutional differences in CCI scores and, from a factor analysis of institutions, obtained four main ways in which environments differed as follows: 1) an intellectual, humanistic, esthetic emphasis; 2) a friendly, group welfare emphasis; 3) a scientific, independent emphasis; and 4) a practical, status-oriented emphasis. In my thinking these four dimensions replaced the 30 environmental press dimensions. Then I also made a content analysis of the 300 items in the CCI to see how well the items reflected characteristics commonly considered in accreditation reviews and college self studies. There were very few items referring to administrative or to academic characteristics and too many items referring to student characteristics. So I developed an instrument in which administrative, academic, and student sources of press were in balance. This questionnaire, called the College Characteristics Analysis (CCA) was also diagnostic of academic and student subcultures in the college. Half of the CCA items were also in the CCI, and half were newly written, for a total of 180 items in the CCA.

At the Behavioral Sciences Center one of the most valuable benefits was the opportunity to have many discussions with scholars from other fields and for me especially with sociologists, anthropologists, political scientists, philosophers, as well as other psychologists and educators. Just as the measurement of individual differences had led to many insights about college students, so now the measurement of college environment differences may lead to further insights about higher education. Before leaving the Center I outlined plans for a ten-year research program.

Back at Syracuse we got CCI responses from more institutions. I received a contract from the Office of Education for a study of academic and student subcultures in college, using the College Characteristics Analysis, which I transferred to UCLA when I changed jobs in the fall of 1961. When I left Syracuse I had available CCI data from about 80 colleges and universities. Since Stern and I were co-authors of the CCI we obviously needed to reach some agreement about who could do what in the future. The result was that Stern would distribute and analyze the CCI as a counterpart to his inventory of personality needs; and I would use half of the original items in an instrument based on environment differences and without regard for the personal needs in Stern's inventory. Subsequently, at UCLA I found that nearly half of the CCI items did not discriminate between environments, so I produced a 150-item questionnaire with the most discriminating items and called it College and University Environment Scales (CUES).

UCLA: ENVIRONMENT, EVALUATION, EFFORT

The professor of higher education at UCLA was Malcolm MacLean, who had been my first employer when he was the director of the General College at Minnesota. His retirement led to my appointment—another instance of my many Minnesota connections over the years. UCLA had a new chancellor, and a dean of Education whose plans for the school emphasized the role of research. I had never before been employed full-time in a School of Education but I knew a few people at UCLA in education and in the psychology department and I accepted the job offer with confidence and pleasure.

Studies of College Environments

For nearly a decade much of my research was devoted to studies of college environments. A factor analysis of responses to CUES from 50 institutions led to the production of five factors—labeled scholarship, awareness, community, propriety, and practicality. These same five factors were always reconfirmed in subsequent analyses. Also, from the beginning, the scoring of CUES was unique. Students' responses were interpreted as a opinion poll. Students were reporters about the environment. When students agreed by a margin of two to one or greater that something was generally true about the college, the statement was counted toward the score. The number of such statements along a particular dimension—scholarship, awareness, etc.—was the score on that dimension. This is not a mean or average score, it is a single score that characterizes the institution. The two-to-one definition was an arbitrary choice. I just regarded it as sufficient to describe something as characteristic or dominant. The difference in result between this method scoring and a score based on the average of student responses is dramatic. For example, if there are 20 statements about the environment and student responses are divided 50-50 about whether each statement is generally true or generally false, the average of student responses would be 10. In the rationale for scoring CUES the institution's score would be zero because no statement was regarded as characteristic by a margin of 2 to 1 or greater among qualified reporters. Also, since there is no mean or variance, the usual methods of estimating reliability are inappropriate. Test-retest reliability can be noted just by seeing whether the college's score is the same on two occasions. It almost always was. Moreover if the first score was 10 and the second score was 11, the second score almost always included the same ten items plus one more. The stability of a score depends on the number of items where the percentage is close to the dividing line of being counted or not counted. CUES was published by Educational Testing Service in 1962.

The research contract to study academic and student subcultures was officially transferred from Syracuse to UCLA as of January 1962, and continued until the final report was issued in 1964. The content of the CCA had three parts, each with its own directions. Part 1, The College or University as a Whole, consists of items that refer to regulations, policies, facilities, and other features of campus-wide relevance. Part 2, Your Major Academic Field, consists of statements about professors, classes. teaching, etc., which the student answers with respect to the academic part of the college he knows best—his own academic major field. Part 3, Your Student Colleagues, is answered with respect to the students and student activities one knows best. Thus the test provides an indication of the source of press (administrative, academic, and student) as well as the direction of press (humanistic, scientific, welfare, practical). Nine institutions participated in the study. Three were small and not likely to have any deviant subcultures; two were somewhat larger; and four were large complex places assumed to have numerous academic and student subcultures. On the back of the CCA answer sheet students indicated their progress toward each of eleven objectives, their grades, and their satisfaction with college. Also at each of the colleges measures of personality were obtained. These differed in the different schools. The measures included the Allport-Vernon-Lindzey values, sections from the Omnibus Personality Inventory, the Heston Personal Adjustment Inventory, Stern's Activity Index, portions of the California Psychological Inventory.

The general purpose of the CCA was diagnostic. In the larger institutions are there many academic subgroups whose environmental press differs significantly from the institution as a whole? Which has the greater impact on attainment—the subgroup or the total group? What happens to similar types of students in different types of subgroups, and to different types of students in similar types of subgroups? In other words, which is more frequently related to attainment—student characteristics or environmental press? From the many analyses made in this research the results indicated that the largest influence on the attainment of general objectives came from the largest stimuli (the total institution), and that subgroups are smaller stimuli having smaller influence. Also differences between institutions were larger than differences between subgroups. The environmental press of a subgroup was more frequently related to attainment than the personal characteristics of students in the subgroup. Viewing each institution as a case study, one could find many interesting differences in the main sources of press—for example, the press toward humanistic objectives might come almost entirely from faculty sources at a particular college, or, at another college the practical, status-oriented press might come mostly from student sources with some added support from administrative sources but no support from faculty sources.

When this study ended, no further research with the CCA was undertaken. I had suggested to ETS that the CCA would be especially well suited for large universities; but ETS decided not to expand its activities in higher education testing and research.

Meanwhile, many more studies of CUES were underway. One series of studies, supported by the College Entrance Examination Board, compared CUES results from different groups of reporters. CUES scores were highly stable over a period of one or two years; almost identical results were obtained from any group of qualified reporters; and results were not influenced or biased by the abilities or the academic success or the personalities of the reporters. Freshmen, however, are not qualified reporters. It may be valuable to have them indicate what they expect will be true about the college and to compare their expectations with what others say is really true.

Over a period of several years, ending with a report in 1967, a grant from the U.S. Office of Education supported many new analyses of the measurement qualities of CUES, a clearer definition of the five factors, the development of a national baseline, the identification of institutional types, and much evidence documenting CUES validity. CUES scores were available from more than 200 colleges and universities. Each school's profile (its five scores) was put into a cluster analysis computer program to locate schools having a similar profile. This was done in more than one way and with more than one sample of schools. The result was always the same—eight types of institutions. The first five types listed below were the clearest; the others were less sharply defined.

SLA Highly selective liberal arts colleges
SU Highly selective universities
ES Colleges emphasizing engineering and science
Den Strongly denominational colleges
TC Teachers colleges
GLA General liberal arts colleges
GU General universities
SC State colleges and other universities

For the national baseline we selected 10 schools of each type except GLA and GU which had 20. Overall the baseline was a mid-way compromise between a sample based on enrollments and a sample based of institutions. The result is a reference group similar in concept to the Dow-Jones index which is not a sample of stocks but rather a selection of stocks in important categories.

For each of the five CUES scores we plotted the distribution for the eight types of schools. This showed clearly that the range of scores within a type was much smaller than the total range, and that in many cases the lowest score of an institution in one type was higher than the highest score in another type. So, there are important differences between institutional types, and that information could be used in college admission decisions if it became available.

With respect to CUES validity, comparisons were made with results from other studies. The National Opinion Research Center had undertaken two large studies of the educational and career plans of college students. More than 300 different schools were involved in those surveys. Responses to CUES had already been obtained at some of the schools. We made an arrangement with NORC to get CUES data from many more of their schools and in return NORC would share with us all of their data. This enabled us to make many connections related to CUES validity. We also had data from the National Merit Scholarship Corporation, from all of Astin's studies, and from CEEB, altogether totalling several hundred correlations comprising a validity network for CUES. The conclusion was that campus atmosphere, as measured by CUES, is a concept buttressed by substantial evidence of concurrent and construct validity.

CUES' second edition, together with a technical manual that included all the new psychometric data, the national baseline, and the validity studies, was published by ETS in 1969.

Evaluation: New Concepts and New Research

In the mid 1960s the U.S.Office of Education decided to fund the establishment of national research centers focused on important educational topics. At UCLA we submitted a proposal for a Center for the Study of Evaluation. With respect to higher education I said we needed a broader range of outcome or criterion measures, and a greater realization of the importance of environmental or contextual measures. When the Center was funded I tried to incorporate these views in the studies of higher education evaluation which I directed for a period of about seven years.

Historically, in education, evaluation was concerned with the evaluation of instruction and courses. One needed specifically defined objectives and specific measures of their attainment. But what if you wanted to evaluate not just a course or a curricula, but the institution, or higher education in the U.S. where there are multiple and conflicting objectives and programs. The familiar experimental models of research and evaluation do not fit this larger problem. A concept of evaluation appropriate for the study of large and complex institutions must ask different questions, proceed in a different style, and have a new view of the purpose of evaluation and the role of the evaluator. In a report I wrote for the Center I summarized this larger concept as follows:

1. It begins with the central question "What are the consequences?" rather than with the more limiting question "What are the objectives?"
2. Its style of inquiry is more aptly characterized by the word "exploration" than by the words "control" and "focus".
3. It sees the role of the evaluator as a social scientist rather than as a teacher, missionary, reformer, or staff officer to the practitioners.
4. Its purpose is to provide more complex bases for informed judgment.

The relevance of these thoughts to some of the current(1996) activities and views about evaluation and assessment will be discussed later in this essay.

The major project undertaken in the Center's evaluation program was, not surprisingly, another questionnaire survey of alumni and current students. In this type of national appraisal there are three potential sources of distortion: first, an inadequate range of criterion variables; second, an inadequate range of contextual or environmental variables; and third, an inadequately representative population of individuals and institutions.

For the population of institutions we selected examples from each of the eight types identified from CUES research. We tried to get 100, as in our national baseline; we succeeded in getting 88. Within each institution, depending on its size, we got random samples of 300 or 150 alumni from the graduation class of June 1950, and similar samples of current freshmen and upperclassmen. The response rate from these groups to the questionnaire was 80 percent from freshmen, 66 percent from upperclassmen, and 58 percent from alumni.

For the criterion variables we had 12 activity scales dealing with a broad range of involvement in contemporary society and culture, a measure of knowledge and awareness about major changes that are taking place in American society and attitudes toward such changes, and ratings of benefit or influence from college toward various educational objectives.

For the environment or contextual variables we had the institutional type from CUES, information about the major field, academic performance, participation in extracurricular activities, aspects of the college experience that stand out in memory, and some corresponding information about high school. Personal characteristics of the respondents are also part of the college context or environment. Beyond the usual census data we included questions about personal and family background, places of residence, personality measures of theoretical orientation, autonomy, and complexity and a vocabulary test.

Most of the results of these questionnaire surveys were reported in my book, *The Demise of Diversity? A Comparative Profile of Eight Types of Institutions*, The Carnegie Commission on Higher Education, 1974. The evidence reconfirms and strengthens the relationships between college environment and experience, on the one hand, and subsequent activities and benefits. And it shows again that there are real differences between some types of institutions, not only in what they offer and emphasize but also in the activities and attainments of their students and alumni.

Unfortunately the full potential of this national study was not realized because the Office of Education terminated the funding. It was the Carnegie Commission, not the federal government, that enabled me to write the book. However, during the years that funding existed for the higher education evaluation program, there were other achievements. This included more than 20 doctoral dissertations that drew upon the data base, or produced additional kinds of measures for future research. In 1971, owing to the government's wishes, we developed a loose-leaf notebook of nearly 40 brief measures, which any college could select for its own particular interests. This included, for example, measures of learning styles, campus morale, teaching, educational preferences, involvement in campus events, societal and educational priorities, etc. More than 1,000 copies of this notebook, called the Higher Education Measurement and Evaluation KIT, were distributed, and several hundred colleges used some of the measures. The KIT also included the various measures from the undergraduate questionnaire—the activity scales, goals, etc.

The decade of the 1970s was a busy one. Clark Kerr asked me to write a book about Protestant colleges. I did, drawing on the recent national survey, CUES data, and some campus visits—*Education and Evangelism,* McGraw-Hill, 1972. I was issue editor for one of the New Directions in Higher Education reports. This was on *Evaluating Learning and Teaching,* Jossey-Bass, 1974. Then there was a FIPSE-sponsored national program called Better Information for Student Choice in which UCLA was one of the participants. For this I wrote a little booklet called *UCLA: Who Goes? What's it Like?* which was widely distributed to high school students, counselors, parents, and others in 1976. In 1978 I wrote a report for the university on what happened to entering students seven years later. And at the end of 1979 Jossey-Bass published a book I had been working on for a couple of years, *Measuring Outcomes of College.*

In 1973 a large national project proposal was suddenly and without explanation not funded. This was to be a study of liberal education in large public universities. For a couple of years, the National Association of State Universities and Land Grant Colleges, together with the Council of Colleges of Arts and Sciences, expressed concern about the status of liberal education in those institutions. They appointed a subcommittee to consider what might be done. The consensus was that in the NASULGC institutions the traditional emphasis on technology and applied science had resulted in both philosophical and financial neglect of the liberal arts, that studies of such trends were needed, and that measures of students progress toward liberal education outcomes were needed. Representatives of the associations initiated discussions with me about my possible interest in developing and carrying out such studies. I was very much interested and said I could devote a substantial portion of my time to it over the next several years. The association then formally invited me to develop a proposal and to solicit funds for carrying it out. There were many discussions with people in the associations and also with Earl McGrath who was serving as a consultant to the Lilly Endowment, and who was especially concerned about liberal education. Together with the association we developed a three-year inquiry for submission to the Lilly Endowment. It was our belief and also Earl McGrath's understanding that the proposal would be presented and recommended for funding at the next meeting of the board of directors. But the proposal was never presented to the board. So that was the end of it. Too bad, because the findings from such a study would be especially valuable today as the content of undergraduate education is being very widely discussed. We know a lot about the effectiveness of selective liberal arts colleges, but we really don't know as much as we should about liberal education in the NASULGC institutions that grant more than a third of all baccalaureate degrees.

Earlier in this section I said we needed to rethink the role of the evaluator. A few years after the UCLA evaluation center was established, the director at the time defined evaluation as follows: Evaluation is the process of determining the kinds of decisions that have to be made—selecting, collecting, and analyzing information needed in making these decisions—and then reporting this information to the appropriate decision makers. This was obviously not my concept. In my view the emphasis on decision makers has several potential dangers. I personally think the evaluator needs to be independent of the decision maker. And who is the decision maker? Is it a professor, a dean, a college president, a board of trustees, a legislature, a governor? Perhaps some distinction needs to be made between an Institutional Research office which obviously provides staff services to the administration, and an Evaluation office where independence is desirable.

I also suggested that the role of the evaluator was that of a social scientist rather than a teacher or reformer. This conflicts with the view of evaluation and assessment as instruments of change and improvement. However, I neither reject nor minimize the role of evaluation as a learning experience leading to changes; I am only saying that for a very large and complex topic, such as higher education in the U.S., or a state system of higher education, it may be desirable to stand back far enough from the activities to view them in a social science research perspective.

The most provocative concept that emerged in my thinking during the 1970s was to question the merit of the input-environment-output model for evaluation and, even more radically, to question the merit of removing student characteristics (input) when judging the influence of environment on outcomes. Before the student arrives at the college, his SAT score is "input"; but after he is in college his SAT score, along with the SAT score of all the other students, is now a characteristic of the student body and is therefore an environment characteristic. If you want to study the influence of the environment why would you want to remove one of the important features of the environment? There are also other grounds for re-considering input defined by student characteristics. The nature of experience itself needs to be considered. From psychological research[1] one can entertain the following thoughts:

- experience consists of events
- events have a quality as a whole
- this quality is the resultant of the experiencer and the world, or physical event
- the meaning of an event therefore consists of the context which the experiencer brings to it, and the context of the physical event

---

[1]See James Jenkins' conceptual explanations of human memory in *American Psychologist*, November, 1974

- the college experience consists of the events that occur in the college environment
- since the experiencer is an integral part of the meaning and quality of an event, the characteristics of the experiencer that are brought to bear on any given event are part of the event itself
- how then can or should the person and the event be separated?

These thoughts form what I have called a contextual model. Instead of input-environment-output, the concept is better expressed by environment-experience-development. Perhaps we should abandon the whole idea of student input. The environment is the input. Its what's there before the student arrives. Then the question is what does the student do in the environment, and how do different students use the opportunities for learning that the environment provides. This line of thinking was one step along the way toward the concept of "quality of effort"

The College Student Experiences Questionnaire

Several past experiences and thoughts contributed to my belief that something called quality of effort could be measured and might add substantially to our understanding of student learning and development. For example, education is both a product and a process. We have typically thought of educational processes in terms of what they contribute to the product; but we know that some processes are qualitatively better than others, just as some products are better than others, so perhaps we should give more thought to measuring the quality of the processes. One motivation for my desire to measure student effort was the recurring rhetoric about accountability that always blamed the institution for outcomes. If students don't graduate it's the college's fault. If they don't learn, it's the teacher's fault. If the graduates don't get good jobs, the college is to blame. If you don't benefit from going to college it's their fault. This assumes that the student is buying a product when actually the student, at a later point in time, is the product. So, the other side of accountability is the quality of effort students invest in using the facilities and opportunities the college provides. Several years earlier in a questionnaire given to UCLA upperclassmen the following statement was included: "If students expect to benefit from what the institution has to offer, they have to take the initiative." Ninety-five percent of them agreed with that statement. They know that they have to make an effort, that it's up to them to take advantage of the facilities and opportunities that are before them.

The model of a Guttman scale seemed to me to be particularly well suited to the idea of measuring quality of effort. A few years earlier I had developed such a scale on the topic of course learning. It occurred to me that the taxonomy of educational objectives developed by Benjamin Bloom and others could be viewed as a hierarchical classification of learning activities as well as a classification of objectives. I developed a scale of learning activities more or less parallel to the levels in the taxonomy, tried it out in a few courses, found that the professors liked it, and that the quality of cognitive effort showed significant differences between A students and C students. The question, then, was whether or not similar scales could be developed for other important aspects of college experience.

Thanks to a grant from the Spencer Foundation I was able to spend two years creating the College Student Experiences Questionnaire (CSEQ). It may be of some value for current researchers and especially those who develop measures to note the decisions and data that ultimately defined the CSEQ. Too often, in my view, questionnaire makers just include items because they are interesting or in some sense relevant to a topic. Given my background, I think of a questionnaire as a test or measure. What is the variable you are trying to measure and how is it defined?

How do you decide what aspects of college life to look at? And how do you decide on the underlying quality dimension of each aspect? There should be some theoretical or conceptual backing for whatever ways one answers those questions. Many college events are related to facilities—classrooms, libraries, etc. These can be seen as behavior settings, for each facility has a purpose and there are characteristic activities that occur in the facility. Major facilities on all or nearly all campuses include classrooms, laboratories, libraries, residences, student unions, athletic and recreation facilities, cultural facilities such as auditoriums, galleries, theater. Other facilities are important on some but not all colleges, such as chapels; and for still others the concept of quality of student effort does not apply such as the health service or counseling office. Other important events are not associated with a specific facility such as a great variety of interpersonal relationships, contacts with faculty members, involvement in clubs and organizations, informal student conversations, etc. So, some scales would refer to students' use of major campus facilities, and other scales would refer to important personal and interpersonal associations on the college campus.

Then I decided that the items in each scale should be simple, direct, clear statements of an activity so that students would know immediately whether they had engaged in the activity and about how often. There would be no statements of opinions or of likes and dislikes or of students' satisfaction or dissatisfaction. Statements of that sort are not indicative of action or effort. I also decided that no statements would refer to events off campus or to facilities which the college itself did not provide since the college is not responsible or accountable for them. Each of these decisions clarifies the definition of effort and rules out the inclusion of unrelated content.

As we started to write statements for the scales we sought advice from many people. We talked to librarians about students' use of the library, with science faculty members about science laboratories, with English professors about writing, with counselors about student acquaintances and experiences that might contribute to self understanding, with student groups in dormitories at UCLA about opportunities for growth and development in that particular setting. Altogether we developed quality of effort scales related to 18 different topics. In some cases we tried out different directions for students' responses—for example, referring to a particular course or to courses in general, or referring to events during the current school year or to events in college up to now. Each scale was printed on a single sheet of paper and we asked friends in various colleges to try out a few of them on a few students. In this way we got from 200 to 500 responses to each scale, from a total of 28 different colleges in all parts of the country.

Now we could examine the statistical properties of each scale. Student answers to each activity were marked by checking *never, occasionally, often*, or *very often*. Admittedly this is not very precise, but the direction of frequency is clear. We had written from 11 to14 items for each scale with the expectation that the final scales would each have about nine or ten items. We scored each scale by giving four points for very often, three points for often, etc. Then we looked at the range of scores for each scale, the item intercorrelations, the correlation between each item and the total score (adjusted), the reliability of each scale, and whether from a factor analysis there was one dominant factor as there should be in a Guttman type of measure. With this data we picked the content for each scale knowing that it was a good measure. The final questionnaire, ready for use, had seven scales related to the use of facilities and seven related to personal and interpersonal events.

The complete printed questionnaire had various items about the students' background and their status in college, their rating of college characteristics along dimensions somewhat similar to CUES, and their estimate of how much gain or progress they felt they had made toward each of a list of important goals. In the spring of 1979 thirteen colleges used the questionnaire—three research universities, three state colleges, two community colleges, and four liberal arts colleges. Now we are ready to find out what we can learn from students' responses to this questionnaire.

What has been learned is a lot. Most of it has been reported in monographs, journals, and in research papers delivered at the national conferences of AERA, AIR, ASHE, NASPA, and AAHE. Following the initial use of the CSEQ in 1979, other colleges heard about it and asked how they could get it. So we set up a distribution office at UCLA. In 1983 I published CSEQ second edition, and in 1990 the third edition. When I left UCLA in 1993 I transferred the CSEQ to Indiana University, and a similar instrument for community colleges (the CCSEQ) to The University of Memphis. There have now been more that 500 colleges and universities that have used the CSEQ. The variable, "student effort", is now a basic element in most research designs for studying student learning and development in college. A few excerpts from my observations along the way probably should be noted.

In my report to the Spencer Foundation in 1979 I ended as follows:

> The most striking findings from this study are the discovery that quality of effort is the most important factor in accounting for students attainment, and that after all other influences have been added together, quality of effort still makes a substantial additional contribution. So it turns out that the most influential variable of all is one that has

not previously been included in studies of higher education. Now that it can be included, we shall see whether new research confirms the conclusion I would draw from the data reported here, a conclusion somewhat at variance with prior research. Put simply, the conclusion is this: What counts most is not who you are or where you are but what you do.

I believe that this conclusion has been consistently supported by the research studies conducted by other investigators as well as by me in the 1980s and 1990s. The concept or variable I have called quality of effort is operationally defined by the activities in the CSEQ, but there has not been a scientific, psychological-educational, dictionary definition. Effort is not the same as motivation. It is not the same as persistence. It is not exactly the same as initiative. It has elements of motivation, persistence, and initiative, but it also has a specific educational context, and its strength probably depends on the context. My general view is that quality of effort describes voluntary behavior. It reflects initiative. It describes the strength and scope of personal investment that students are making for their own higher education.

Based on the CSEQ responses of 25,427 students from 74 colleges and universities obtained during the years 1983 to 1986, I wrote a monograph, *The Undergraduates,* UCLA Center for the Study of Evaluation, 1990. It is not a technical monograph. There are no means and sigmas, no significance tests, no correlations, no regressions—just percentages and averages and descriptions of what students reported about their activities and progress. I wrote it that way because I hoped that nonspecialists could read it, especially perhaps some of the people who had read the critical reports that had been recently published. One of those reports criticized the curriculum, another one criticized the professors, another criticized the lack of evaluation and measurement, and another criticized the administration. None of the reports dealt directly with the students' experience in higher education. I think the writers just assumed the students were not learning much. From the CSEQ data I came to very different conclusions:

> We found, contrary to the critics who claim that students don't learn anything, that all students believe they have made at least some progress toward every one of the important goals listed in the questionnaire. What is even more dramatically contrary to the critics is the finding that a majority of students at all types of institutions believe they have made substantial progress toward ten of the most basic and historically most highly regarded goals of higher education—gains in intellectual skills, in breadth of knowledge, in vocational preparation, and in personal and social development. If we are to believe the students, the critics are wrong and badly misinformed.

We also found that the "average" student spends about 35 to 40 hours a week on academic activities. In some places, especially in the selective liberal arts colleges, many students spend 50 hours a week or more.

> The level of effort and the scope or breadth of effort students put into their college experience was found to be a very good indicator of the quality of the undergraduate experience, clearly associated with progress toward all important goals, with better grades, and with greater satisfaction.

One of the important findings in this national survey is the truly distinctive character of the selective liberal arts colleges. They are uniquely powerful environments for student learning and development. Nationally, the Carnegie Foundation identifies about 125 selective liberal arts colleges, and I estimate that the total number of undergraduates enrolled in them is roughly 3 percent of the national total. Are they an endangered species?

VIEWS, VIRTUES, AND VALUES

For about a century the trend in the shape of higher education in the U S. has been away from liberal arts colleges to large comprehensive universities, and from private to public institutions. In the last few decades there has also been a trend toward greater enrollment of nontraditional students—older, part-time, nonwhite. And most recently there is exploration of higher education via computers, WWW, Internet, etc. All this is relevant to understanding the results of nearly all of our surveys because nearly all of the students who answer our questionnaires are traditional full-time students on the campus. We know that nontraditional students are less likely to answer questionnaires; and we need to figure out better ways to reach them and perhaps also better questions to ask them. None of this invalidates the results of our past surveys; it just says that there are other populations and conditions we need to consider in the future. The college campus, with full-time students living away from home, is the hard core of higher education. Nearly all liberal arts colleges are in small towns. And many of the major universities are not located in population centers and do not have part-time commuting students—Princeton, Dartmouth, Cornell, Duke, Georgia, Michigan State, Notre Dame, Indiana, Purdue, Iowa, Louisiana State, Stanford, and many others. So for millions of students going to college means going to a particular place and living there; and the college environment is still an influential variable. Higher education is both an individual and a social experience. This personal/social combination, in behavior settings designed to facilitate learning and development, is a memorable and influential experience. "Virtual experience," via the best of modern technology, is not a substitute for real experience. And part-time education is not the same as full-time education. Whether any of this matters very much depends on what we think higher education is, or should be.

At the AIR Forum in the spring of 1996 I heard a speaker who thought of higher education as a commodity with its condition and value determined by the market. This reminded me to look back at a speech I gave in Iowa twenty-five years ago. Here are some excerpts:

> Many of the analogies and models we use in thinking about education and evaluation are drawn from fields that have no necessary connection with the nature and quality of education. Higher education is not a factory, receiving raw material, processing

> it, and turning out products having certain performance characteristics. Nor is higher education a business, distributing goods at so much per unit cost. Nor is it a bureaucracy run by bosses with flow charts, communication networks, decision points, and job descriptions. But these mechanical and administrative analogies have their counterparts in the language of educational research and evaluation—the measurement of input-output differences, specified performance objectives, college effects, test score gains, etc. Most recently the popular terminology includes behavioral objectives, product development, cost effectiveness, performance contract, management information systems, and accountability....These conceptualizations tend to emphasize and reinforce an administrative and efficiency view of the nature and purpose of education....A college should be judged by the quality of life that it fosters, the opportunities for experience and exploration it provides, the concern for growth, for enrichment, and for culture that it exemplifies....College could be conceptualized as an environment for exploration—of self, of knowledge and skills, of ideas and values, of society, conscience, community, and commitment.

The variables and concepts that are relevant and useful in other disciplines are not necessarily relevant or helpful in the study of higher education. In any case, it is important to realize what image about the enterprise is guiding what we see and say about it, for all thought begins with an image or at least reflects one.

To a large extent the value of our research can be judged by what we look for and how adequately we measure it. From the General College at Minnesota I saw the value of knowledge about college graduates and the scope of what could be learned from a questionnaire. Later I constructed a set of reliable and valid scales to measure the quality of life of college alumni—responsible citizenship, cultural participation, breadth of engagement in adult society. Next, from my association with the College Entrance Examination Board, and the Social Science Research Council, I sensed the value of knowledge about the college environment. This led to the construction of CUES, enabling the quality of the institution to be measured by the collective perceptions of its students. The characteristics of an institution could then be seen as an educational/psychological stimulus or context for student learning and development. Then, from an interest in the evaluation of teaching and learning, and a concern about institutional accountability, I created a way to measure the quality of effort students invest in their own education. I think each of these lines of inquiry has contributed to the enlargement of research in higher education—the quality of life among alumni, the quality of college environments, and the quality of effort among students.

From my experiences as a researcher, a teacher, and an administrator, I see higher education as a field of study, an area of inquiry, not as an academic discipline or specialization. The best background for a college president is a knowledge of history and philosophy—an understanding of how the college came to be what it now is, and what values are crucial in guiding its future. In the more limited field of measurement and evaluation I think the basic requirements for good evaluation are judgment and creativity, and the basic requirements for writing good test items are knowledge of the subject and English composition. Beyond that, various psychometric methods and research designs come into play. The danger in thinking of higher education as a narrow specialty is the likelihood that one's research will focus on limited and easily defined topics, and will be another example of what I once described as unsuccessful efforts to find large significance in small phenomena. In the physical and biological sciences the exploration of ever smaller units of analysis has resulted in new insights—a particular gene, a DNA trace, atoms, quarks, etc. I'm not at all sure that the domain of the social sciences is similar. New insights in the social sciences may come from seeking to discover larger combinations of evidence, and understanding the way events and conditions fit together to create a major influence. If so, progress will come from expansion, not from reduction.

Also, we need to be reminded from time to time that the value of results from a questionnaire survey or a test does not depend primarily on the quantity of responses; it depends on the quality of the questions. The emphasis on quantity as in collecting large national samples usually produces very many very low correlations that are nevertheless statistically significant. In some reports I've read, correlations of .10 are significant and also thought to be important. If we recall the old "coefficient of alienation" we would realize that a correlation as high as .30 still leaves 91 percent of the variance unaccounted for! Often too, large national surveys do not have adequate replies from single colleges; yet if changes are desirable they have to be made at the local level. Because large scale surveys attract attention, the quality of the questions is especially important. Most large scale surveys do not consist of carefully developed measures of fundamental constructs. What are the basic variables being measured? and with what validity? The big surveys "cover" various topics, but not fundamental concepts. All this gets us back to thinking about our purposes and programs.

My general interests and inclinations have been to look for relationships in their natural setting—between environments and attainments, between effort and outcomes, between the pattern of activities as college students and the pattern of activities as college alumni, between institutional purposes and institutional influences. I believe that making connections is a good way to stimulate new research and new insight. But who knows what connections will be important in the future? The size and shape of higher education as we know it today may be very different tomorrow. In any case, I am sure that higher education will continue to be a fascinating topic of research.

Reflections on the Study of Effective College Teaching and Student Ratings: One Continuing Quest and Two Unresolved Issues[1]

Kenneth A. Feldman
State University of New York at Stony Brook

When I was invited to address the Special Interest Group on Faculty Evaluation and Instruction of the American Educational Research Association (Feldman, 1995), I must admit that, initially, I was not altogether sanguine about preparing a presentation. After years of reading and integrating the research on instructional effectiveness in higher education, particularly the research on college students' assessment of their teachers, I felt that most (if not all) of what I had to say or report I had indeed said or reported—in a series of research integrations (Feldman, 1976a, 1976b, 1977, 1978, 1979, 1983, 1984, 1986, 1987, 1988, 1989a, 1989b, 1990a, 1992, 1993; also see Feldman and Newcomb, 1969, Chap. 8) and in some occasional pieces (Feldman, 1973, 1990b, 1994). So I was concerned about merely repeating past observations.

Yet, as I reflected on the work I had done over the years, I decided that there might be some value in highlighting a continuing research quest in the field—that of establishing the conditions and contexts that determine the strength and patterning of pertinent empirical associations. It also seemed worthwhile to elaborate on two long-standing issues in the field that have never been fully resolved: one dealing with the question of bias in college students' ratings of teachers; and the other concerning the applicability of the traditional model of psychological

---

[1]This chapter is based on an invited address (in conjunction with receiving the Wilbert J. McKeachie Career Achievement Award) that I presented to the Special Interest Group on Faculty Evaluation and Development at the 1995 Annual Meeting of the American Educational Research Association (Feldman, 1995). Raymond Perry, an Associate Editor of this Handbook, has for some time been encouraging me to bring together into one piece some of the different parts of my work on teaching effectiveness and student ratings; I wish to thank him for this support and for his most helpful suggestions on drafts of this chapter. William Cashin and Herbert Marsh—the two "official" reviewers of this chapter—made thoughtful, penetrating comments on an earlier version. I also want to thank the following persons for their valuable suggestions: Philip Abrami; Anthony Greenwald; Wilbert McKeachie; Harry Murray; and Harry Tagomori.

J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*

testing to student ratings. The quest will remain as long as the field does, and the two issues can hardly be resolved in this chapter. Still, clarification of the "problems" involved should be of some use to the study of instructional effectiveness and student ratings of teachers.

ESTABLISHING THE CONDITIONS AND CONTEXTS THAT DETERMINE THE STRENGTH, DIRECTION, AND PATTERNING OF RELATIONSHIPS

The continuing quest is one that all fields of social and behavioral sciences face—namely that of establishing the conditions or contexts under which relationships become stronger or weaker (or nonexistent), or change in some other way (reversing direction, becoming curvilinear rather than linear, and so forth). More or less equivalently put, the quest calls attention to the importance of determining "interaction effects" as well as "main effects." Establishing conditions and contexts (interaction effects) is important for empirical associations (main effects) discovered in individual pieces of research as well as for any relationships found in meta-analyses and research integrations.

I will give three examples of what I have in mind, primarily based on findings from my own research syntheses. These examples concern the association between research productivity and instructional effectiveness, between specific instructional practices and overall effectiveness of the teacher, and between the gender of the teacher and his or her ratings by students. I have chosen these three areas of discussion not only because of their intrinsic interest but also because some work in establishing conditions or contexts has already begun in each of them. In the discussions of each of these three areas, I will assume that student ratings are valid indicators of teaching effectiveness. Certain questions relevant to the validity of these ratings will then be raised in later sections of the chapter.

The Association Between Research Productivity and Instructional Effectiveness

In one publication (Feldman, 1987) I reviewed the research on the connection between research productivity or scholarly accomplishment of faculty members and their overall teaching effectiveness (as assessed by their students). Across some 29 studies in which the association between the research and teaching variables were reported in the form of product-moment correlations or where the results could be converted to such correlations, the average correlation was +.12. (For a similarly small correlation, based on a recent meta-analysis, see Hattie and Marsh, 1996.) Both the existence and strength of the relationship between research productivity and teaching effectiveness no doubt vary with circumstances or conditions. Indeed, there may be discoverable contexts in which positive associations can be expected routinely to occur, or to be larger rather than smaller.

As an example, a little evidence exists that positive associations between research productivity and instructional effectiveness are more likely to occur, and to be larger, in the humanities and the social sciences. As a possible explanation, Michalak and Friedrich (1981) have hypothesized that "research in the natural sciences, in contrast to research in the social sciences and humanities, may be at a level of abstraction and complexity that renders it of little utility in the [undergraduate] classroom" (p. 593). Note, however, that the conclusion of greater associations between research productivity and teaching effectiveness in the social sciences and humanities than the natural sciences is extremely tentative—based as it is on only five studies with appropriate data (see Feldman, 1987, pp. 269-273). Because of the nature of its data, findings for a study by Hoyt and Spangler (1976) could not be included with the other five studies in the meta-analysis under consideration. Were these results able to be included with the other five studies (and given certain assumptions), it is quite possible that humanities would have the highest association between research productivity and teaching effectiveness, followed by the social sciences and the natural sciences (with some possibility that the associations for these latter two would end up not far apart in size). If so, this pattern of results would be congruent with the explanation that the association between the two variables under consideration would be expected to be greater in the humanities where originality lies mainly in creative scholarship than in the natural and social sciences where originality lies in experimental and empirical research (see Ben-David, 1977, and Elton, 1986; but see Hattie and Marsh, 1996).

Evidence is accumulating that academic departments and divisions differ in their faculty's attitudes toward students, classroom teaching practices, instructional and educational goals, and other preferences and beliefs, although these differences are not always large (Blackburn and Lawrence, 1995; Braxton and Hargens, 1996; Hativa and Marincovich, 1995; Stark and Lattuca, 1997). Such differences in faculty attitudes and behaviors—in addition to (or instead of) disciplinary differences in levels of abstraction or complexity and bases of originality—might account for any disciplinary differences in the association between research productivity and teaching effectiveness. Philip Abrami (personal communication, February 27, 1996) further suggests that if differences exist across disciplines in time spent by faculty members on teaching and research, and in their perception of how efforts spent on teaching relate to student learning and their own career success, then these particular differences in whole or in part might also mediate the association between research and productivity and teaching effectiveness. Whether any of these speculations are, in fact, true awaits further research. (For other possible "moderators" of the association between research and teaching, see Braxton, 1996, and Hattie and Marsh, 1996. Both of these valuable publications became available too late for full incorporation into the present analysis.)

It is also conceivable that two, three, or even more contexts or conditions may combine to produce higher-level interaction effects (cf. Centra, 1983). Perhaps for certain departments, academic divisions or disciplines at certain schools (and even then perhaps only for faculty at certain career stages), research productivity may promote much larger positive relationships between research and teaching than are generally found. What these specific conditions or contexts are, or whether they even exist in reliable and specifiable ways, is currently unknown.

Gage (1991) has pointed out that one implication of the low correlation between research productivity and teaching effectiveness is that faculty members can be found in all four quadrants of a scatterplot between the two variables: high in both research productivity and teaching effectiveness, low in both, high-low, and low-high. As a consequence, one could investigate the average or general differences between faculty in two or more of the different "groups" in terms of the beliefs, dispositions, and actions of these faculty, as well as possible differences in their circumstances. Although I know of no study that has used exactly this procedure, a study by Kremer (1991) did explore the differences among faculty who were "high" or "low" not only on teaching effectiveness and on research productivity but also on service activities. Using three areas of performance allows in principle the comparison among eight types of faculty: the "all stars" (who are high in performance in all three areas); the researchers and teachers (high on research productivity and teaching effectiveness); teachers and good citizens (high on teaching effectiveness and service); researchers and good citizens (high on research productivity and service); researchers (only); teachers (only); good citizens(only); and the uninvolved (low performing in all three areas). In actuality, only five types were found at the school studied by Kremer (1991); researchers and teachers, researchers and good citizens, and good citizens were *not* found as distinct types.

One especially interesting finding that emerged from Kremer's research (1991) was that associate professors were more likely to be "all stars" than were assistant or full professors. Kremer speculates that as faculty progress from assistant through associate professorships, they may tend to expand the focus of their efforts from strictly research to a combination of research, teaching and service—possibly to develop new interests and balance in their academic concerns (perhaps to avoid burnout and stagnation). Full professors may not feel the same pressures to excel in all academic pursuits and thus may return to being more focused in a particular area, emphasizing teaching or research or service, or simply devoting less time and energy to their academic career.

The Association Between Specific Dimensions of Instruction and Overall Effectiveness as a Teacher

A second example of exploring the conditions or contexts that affect relationships has to do with teaching practices that affect student learning. Although various dimensions of instruction are important to effective teaching, one would assume that some of them are more important than others. One way of establishing this differential importance is to determine how various teaching dimensions relate to student learning, which Cohen (1980, 1981, 1987) did in his well-known meta-analytic studies of the relationships of student achievement with eight different instructional dimensions. Based in large part on work by Abrami and d'Apollonia as well as Cohen himself (Abrami, Cohen and d'Apollonia, 1988; d'Apollonia and Abrami, 1987, 1988), I extended Cohen's meta-analysis by using less heterogeneous categories for coding the evaluation items and scales in the studies under review, widening the range of instructional dimensions under consideration, and preserving more of the information in the studies Cohen used in his meta-analysis (Feldman, 1989b, 1990a; also see Feldman, 1996, 1997)[2]. To be included in Cohen's meta-analysis or my own, a study had to provide data for actual college classes; use the class or instructor, not the individual student, as the unit of analysis; be based on a multi-section course with a common achievement measure for all sections (usually an end-of-course exam); and provide data from which a rating/achievement correlation had been or could be calculated.

The size of the correlation of student ratings with achievement did vary by the instructional dimension being rated. To give only a few of the findings: The highest two correlations of .57 and .56 were found between student achievement and ratings on the teacher's preparation and organization of the course, and the teacher's clarity and understandableness, respectively. A smaller correlation was found for instructional items dealing with the teacher's encouragement of discussion and his or her openness to the opinions of others ($r = .36$), and a still smaller one for evaluation items dealing with the teacher's or course's intellectual challenge and encouragement of independent thought ($r = .25$).

These results might well be different under different conditions or for different contexts. Most of the multi-section courses in the studies under review were introductory courses or lower-level courses (introductory psychology, introductory economics, introductory chemistry or some other natural or physical science, lower-level courses in mathematics, beginning courses in languages, introductory courses in communications, speech courses, etc.). Moreover, achievement tests of the kinds used in these studies, as McKeachie (1987) and Cashin, Downey, and Sixbury (1994) have pointed out, typically measure lower-level cognitive outcomes such as memory of facts and definitions, rather than higher-level outcomes such as critical thinking or problem solving. Thus the importance of various instructional dimensions might well shift when consideration is of more advanced courses and for higher-level cognitive outcomes—under such circumstances, encouragement of class discussion and intellectual challenging students

[2]Additional analyses and explorations of data from multisection validity studies can be found in Abrami, d'Apollonia, and Cohen (1990), Abrami, d'Apollonia, and Rosenfield (1996), and d'Apollonia and Abrami(1996). Some disagreement exists over the strengths of the multisection validity design and how much reliance is best placed on multisection validity studies—for example, compare Marsh and Dunkin (1992, pp.169-171) with Abrami, d'Apollonia, and Rosenfield (1996, pp. 213-234).

might well be more important to effective teaching. Cashin, Downey, and Sixbury (1994) put it this way:

> ...multisection courses are primarily freshman courses and are therefore likely to have course goals at the lower end of Bloom's taxonomy, where the instructor's organization is important for helping students learn facts. However, if an instructor is teaching higher level thinking skills, then more interactive learning methods would be more likely to have the higher correlations (p. 650).

I know of no study that directly addresses (in all particulars) the proposition that the importance of interactive learning methods, encouragement of class discussion, and intellectually challenging of students become greater than that of preparation/organization and clarity/understandableness when shifting from lower-level to higher-level cognitive outcomes. For reviews and discussions of studies with partially relevant (or indirectly relevant) data, see Pascarella and Terenzini (1991, Chap. 4) and McKeachie et al. (1990, Chap. 3); also see Pascarella et al. (1996).

Informed speculations have been offered as to why correlations are large or small for particular dimensions and what conditions might change the magnitude of these correlations (although empirical substantiation of these conjectures is still needed). For instance, McKeachie (1997), noting that the instructional dimension of feedback from teachers to students does not on average correlate particularly well with student achievement (r = +.23), points out that feedback can have unintended effects dependent on the context on which it is given and the particular meaning it has for students. For example, a student can take criticism as evidence either that one lacks the ability to succeed or, to the contrary, that one actually has the ability to improve. Thus, both the kind of feedback and the relationship between the student and teacher can determine whether the feedback produces a reduction or an increase in motivation.

Gender of the Teacher and Student Evaluations

As a final example, consider the importance of establishing conditions and contexts in determining whether the gender of the teacher is associated with student evaluations of the teacher's overall effectiveness. In a relatively recent research integration of relevant studies (Feldman, 1993), I reported that the majority of studies did not show a statistically significant difference between the two genders in the overall evaluations they received from their students. Where statistically significant differences were found, more of them favored women than favored men. Taking all studies showing the association between gender and overall ratings by a correlation coefficient (whether or not statistically significant) or having data from which this coefficient could be derived, the average correlation also showed women to have higher overall evaluation than men. However, the average correlation (r = +.02) was so small as to be substantively negligible even though it is statistically significant based on combining individual probabilities.

Because various characteristics of the teachers themselves, of students, and of the situation may be associated with both the gender of the teacher and his or her overall evaluation, studies finding (and sometimes even those not finding) an association between gender and overall evaluation have controlled for one or more of these pertinent factors. When a zero-order relationship is originally found between gender and overall evaluation, one approach is to see whether the association remains the same across various conditions or contexts, or whether it is somehow modified (either in its strength or its direction) in certain of the conditions or contexts, or whether it disappears altogether under controls. For example, if it is found that female teachers receive a somewhat higher global rating than male teachers, one might want to see if this association remains the same in different academic areas (natural sciences, social sciences, humanities, etc.), or is particularly strong for certain of the academic areas and weak in others, or disappears within each of the academic areas.

The major control variables used in extant research have been: gender of the student; academic rank of the teacher; academic field or disciplinary area; class level of the course; difficulty of the teacher and course; and the teacher's personality characteristics and orientation to teaching. Establishing the conditions or contexts under which associations are stronger or weaker (or nonexistent) has generally been done by searching for statistically significant interaction effects. Although individual studies have found such interaction effects, these effects appear to be relatively consistent across studies for student gender only. Here, a tendency toward same-gender favorability has been found across studies: in studies with pertinent data, students tended to rate same-gender teachers a little higher than opposite-gendered teachers (see Feldman, 1993, pp. 167-170). What is generally unknown from existing studies, however, is how much of this result is due to male and female students taking different classes or majoring in different fields (and thus having different teachers) and how much is due to differences in the preferences of male and female students within classes. Studies having information about interaction effects (or lack of them) with other aspects of the situation, of the teacher, and of the students (other than gender) do not reveal uniform or consistent interaction effects for any of these aspects. As a simple example, in a study by Aleamoni (1978) of students in anthropology courses male and female teachers were rated differently only in freshmen-level courses (in this case male instructors were rated somewhat higher). By contrast, in a study by Wigington, Tollefson, and Rodriguez (1989) of a broad range of students, it was in upper-division courses where gender differences materialized (again, male instructors were rated higher).

Even when the various attributes, behaviors, and "positions" of teachers (e.g., their academic areas) are grouped together into those that are considered to be gender-typical and those considered to be gender-atypical, consistent interactions effects are not found. That is, sometimes gender-typical attributes, behaviors, and positions of teachers "enhanced" their ratings whereas sometimes gender-atypical attributes, behaviors, and positions did so. The fact of such inconsistencies does not argue for abandoning the search for conditions and contexts, but rather for establishing more complex ones—in essence, looking for higher-order interactions. That is to say, the research quest becomes finding the conditions under which gender-typical attributes, behaviors, and positions "enhance" student ratings and the conditions under which gender-atypical attributes, behaviors, and positions do so.

Additional Comments on Conditions and Contexts

The three examples that have been given hardly exhaust the areas in which the directions, patterns, and magnitudes of associations are potentially affected by discoverable factors that impinge on them. A few other areas of interest from my own work over the years for which more research is needed on contexts and conditions include: the effect of class size on student ratings of teachers (Feldman, 1984); the relationship between teachers' personality characteristics and attitudes (as distinguished from their pedagogical practices) and instructional effectiveness (Feldman, 1986); the connection between seniority or experience of teachers and their perceived effectiveness in the classroom (Feldman, 1983); and the extent of the similarity (or dissimilarity) between what students and teachers believe is important to good teaching (Feldman, 1988). It is true, I might note, that even if individual studies themselves do not establish conditions or contexts, some of them nevertheless can be determined by procedures used in meta-analysis and research integrations. However, the more that conditions and contexts are studied in individual pieces of research, the easier and better they can be established in meta-analyses and research integrations.

Work in the areas under consideration need not—in fact, should not—be limited to student ratings (as indicators of teacher effectiveness) and the factors that are correlated with these ratings. Although not a focus of the present chapter, observational studies of college teachers in the classroom (using trained observers) and experimental manipulation of pertinent factors in controlled laboratory settings are two other important ways to establish conditions under which teachers are more rather than less effective (for examples and reviews of such research, see Murray, 1991, and Perry, 1991). Of the value of experimental research on effective instruction, Perry (1991) has written:

> In the experimental approach, instruction is deemed to be an independent variable having direct effects on specific academic outcomes. It can be operationally defined as an independent variable either through classification procedures or through direct manipulation. Thus, levels of instruction can be identified by selecting instructors who represent a given level, or by training instructors to enact that level. This approach has a distinct advantage over the others because it enables cause-and-effect relationships to be determined and because it provides a more fine-grained account of teaching-learning processes (p. 21).

Finally, not to be overlooked in establishing conditions and contexts of effective teaching is the use of survey questionnaires or in-depth interviews of the parties involved, field experiments, ethnographic accounts, and case studies (Paulsen and Feldman, 1995).

In principle, and clearly in practice, the search for the conditions and contexts that determine the existence, strength, direction, and pattern of associations between variables of interest is an on-going search and probably a never-ending one. Such searching is the very life blood of the social and behavioral sciences. In contrast, there are issues in the study of student ratings and the effectiveness of college teachers that are potentially resolvable. (Whether such issues ever will be resolved is another matter, of course.) I am interested here in two of these issues—two *sets* of issues really. The first set (about questions of bias) can be presented more briefly than the second set (about the conceptual and measurement model underlying the use of student ratings).

MEAN RATINGS OF COURSES AND TEACHERS—QUESTIONS OF BIAS

One set of issues arises when exploring possible biases in student ratings (Abrami, d'Apollonia, and Rosenfield, 1996; Feldman, 1984, 1993, 1997; Marsh, 1984, 1987; Marsh and Dunkin, 1992). I focus here on the *group* level of analysis—that is, on average (or aggregate) ratings of courses and teachers—and the attributes of *teachers* and *courses* that might create bias in these ratings. At the moment there is no clear consensus on a definition or conceptualization of bias in this area of student ratings, although there is general agreement that "the mere existence of a significant correlation between students' evaluations and some background characteristics [characteristics of the students, teacher or course, and situational circumstances] should not be interpreted as support for a bias hypothesis" (Marsh, 1987, p. 310). I take bias to refer to one or more factors directly and somehow inappropriately influencing students' judgments about and evaluation of teachers or courses. This approach is consistent with Marsh's conceptualization of bias: "It is not sufficient to show that some variable is correlated with student ratings and that a causal interpretation is warranted; it must also be shown that the variable is *not* correlated with effective teaching" (Marsh, 1987, p. 310). In this respect, bias means something other than (or more than) the fact that student ratings may be influenced by conditions not under the teacher's control or that conditions may somehow be "unfair" to the instructor (making it harder for him or her to teach well and thus to get high ratings compared to teachers in "easier" pedagogical situations).

To give an example, determining whether the teacher being a man or woman biases ratings by students goes hand in hand with determining whether students ratings are equally valid indicators of teaching effectiveness for both male and female teachers. To be unbiased (and valid) for male and female teachers, the evaluations should not be determined in whole or in part by the mere fact of the teacher's gender—to which students react positively or negatively as the case may be—but should be determined by the teacher's actual dispositions, practices, and effectiveness in the classroom. Incidentally, Abrami (1989) reminds us that "the converse of Marsh's definition is also true; a bias also exists when some variable is both theoretically and empirically related to effective teaching measures but is *not* correlated with ratings" (p. 45, emphasis in original). This would be the case, for instance, if female teachers at a particular school on the average were actually more effective teachers than male teachers, but did not receive higher average student ratings.

If a condition or factor, then, actually affects teachers and their instruction, which in turn is accurately reflected in students' evaluations, bias is *not* present. By contrast, if in some way this condition or factor only affects students' attitudes toward the course and students' perceptions of instructors (and their teaching) such that the evaluations do not accurately reflect the quality or effectiveness of the instruction that students receive, bias *is* present. To give a simple case in point, teachers in large classes may receive slightly lower ratings because they indeed are somewhat less effective in larger classes than they are in smaller classes, not because students, say, take out their dislike of large classes by rating the teachers a little lower than they otherwise would. So, while it may be somewhat "unfair" to compare teachers in classes of widely different sizes, in this particular example the unfairness lies in the difference in teaching conditions, not in a rating bias as defined here.[3]

As a second example, consider Franklin and Theall's (1993) finding of a significant difference (.3 in mean scores on overall ratings) for male and female faculty in a particular discipline at a large multidisciplinary research university. These researchers felt that the differences could have been due to "teaching assignments and a departmental dynamic which put female faculty at a disadvantage" (Franklin and Theall, 1995, p. 45). Assuming that their surmise is correct, the student ratings accurately reflected departmental practices that were unfair to female teachers, but by the conceptualization offered here the ratings were not biased per se.

Other conceptualizations of bias have been offered. At one point, for instance, Cashin (1988) wrote:

> Marsh suggests that student ratings are biased only to the extent that they are influenced by variables *not related to teaching effectiveness*. When using ratings for personnel decisions or the instructor's improvement, I would suggest an even narrower definition, restricting bias to variables *not* a function of the *instructor's* teaching effec-

---

[3]Marsh and Dunkin (1992) review some evidence showing that "the effect of the teacher on SETs [student evaluations of teaching] is much larger than the effect of the course being taught" (p. 160). If the findings they review generally hold across a variety of courses and colleges, one would expect that the effects of various conditions or factors would be "minimized" although most probably not eliminated outright.

> tiveness. Student motivation or class size may impact teaching effectiveness, but instructors should not be faulted if they are less effective teaching large classes of unmotivated students than their colleagues are with small classes of motivated students. In this case, student motivation and class size, although they are related to teaching effectiveness, are not a function of the instructor's characteristics, but of student and course characteristics, and so they should be considered sources of bias (p. 3, emphasis in original).

Because these are definitional and conceptual issues, it is hard to say at this point that one view is better than the other, but it has to be confusing to the field for it to have incompatible views. Indeed, Cashin (1995) himself has come to agree, and has recently written:

> Feldman...[has] observed—accurately in my judgment—that such a definition of bias [as that given in Cashin (1988)], while possibly acceptable, was not the usual definition and it served to confuse the literature. Marsh and Dunkin (1992)—considering that prior student interest in the subject matter is *not* a bias because it does impact teaching and learning—raise the question of "fairness" in comparing instructors teaching classes of interested students versus instructors teaching classes of uninterested students.
>
> In the interest of clarity, rather than using "bias" in the restricted sense I did in the original paper [Cashin, 1988], I will identify variables (when correlated with student ratings) that require control, especially when making personnel decisions (p. 4).

Whether for reasons of reducing bias or making comparisons among teachers fairer, procedures for controlling pertinent characteristics are available. Some of these procedures are relatively complex. For example, Rose (1975) advises using data from multiple regression analyses to actually adjust student ratings. Dilts (1980) uses regression analyses to establish tables of "neutral estimation values" and "neutrality ranges." Haladyna and Hess (1994) suggest the use of a many-faceted Rasch model. A simpler procedure is to form comparison groups (sometimes called normative groups) that are homogeneous with respect to potential biases—for example, establishing the average ratings of teachers of small classes and of large classes (or better, small, medium-sized, and large classes). Thus, if one were teaching a large course, one could compare one's ratings with those of other teachers of large courses.

Controlling for biases (or to make comparisons among teachers fairer), however, has its own issues or problems. Perhaps the most important one is what might be called "overcontrolling"—that is, removing actual teacher effects in the control procedures. Marsh (1987), for example, notes that a control procedure would not be appropriate: "...if inexperienced, less able teachers are systematically assigned to teach large introductory classes, then statistically removing the effect of class size is not appropriate" (p. 311). He does not say why it is not appropriate, but I assume what he has in mind is that controlling for class size in this case is controlling for actual teacher effects (rather than merely controlling, say, for a bias factor). Abrami (1993) has said this more directly: "The field user [of student ratings] faces the dilemma of forming norm [comparison] groups to control for class size effects while trying to insure against removing any instructor effects which happen to covary with class size" (p. 8).

We need to know more about which teachers get assigned to which sorts of courses—courses that themselves may vary in the intrinsic ease or difficulty of teaching. We also need to know which of these assignment practices are local (that is, particular to the college) and which seem to generalize across a wider number of colleges. Without this knowledge, we cannot be altogether certain that controlling for what we think may be a biasing factor (or some other circumstance that potentially produce unfair comparison among teachers) may also be controlling for actual teacher effects.

Abrami (1993) has pointed out that in trying to be fair to teachers by establishing specific norm groups when judging their teaching competence on the basis of student ratings, unfairness can be introduced in certain situations. He gives the following example:

> Imagine separate norm groups formed on the basis of course level (upper level vs. lower level). Here, our hypothetical promotion and tenure committee might argue that students in upper level courses are brighter and more highly motivated than students in lower level courses, and therefore are more likely to assign higher ratings. In addition, upper level courses are usually smaller in size, less likely to be required courses, and may have less stringent grading standards. Again, one would suspect that the instructors of upper level courses seem predestined to receive higher ratings than those of lower level courses. Now consider the type of instructor likely to be teaching these upper level courses. In many instances, the upper level courses are "prized" course[s] that tend to be assigned to more experienced faculty or to faculty with established reputations as good teachers. In other words, there is the real possibility that those teaching upper-level courses will tend to be the better instructors, whereas those teaching lower-level courses are the poorer instructors.
>
> What of [two hypothetical teachers] Smith and Jones? Professor Jones, who is unlucky enough to be teaching upper level courses, will have his student ratings compared to those of the department's better teachers. Because his teaching is not outstanding, Jones's ratings are worse than the average upper-level instructor and the committee reaches a negative decision about his teaching ability. Professor Smith, on the other hand, teaches lower level courses and is lucky enough to have his student ratings compared to those of the department's poorer teachers. Professor Smith receives a favourable vote from the promotion and tenure committee... (p. 7).

Abrami (1993) puts the point in a general way, as follows:

> ...establishing standards of instructional effectiveness based on norm groups may be problematic, leading to inappropriate judgements of teaching competence. One reason is that it is often necessary to make a critical, perhaps untenable, assumption that the average quality of instruction across norm groups is constant [that is, does not vary] and, therefore, rating differences between groups must be a function of student and course characteristics alone (p. 5).

Again, knowledge of which sorts of teachers get assigned to which sorts of courses is essential, as Cashin (1994) points out in his response to Abrami (and note Cashin's "twist" on the idea of "fairness"):

> Abrami [1993] details a scenario which suggests that the higher ratings in upper level courses may be because "better" teachers are assigned to them—a reasonable hypothesis in my judgment—and not, say, because the students were more motivated. Thus, Prof. Jones (described as a not outstanding teacher) might be unfairly evaluated because he was being compared against an above average group. It seems to me that Plan A would be to test whether "better" teachers are assigned upper level courses. However, if they are, is it unfair to expect faculty teaching such course[s] to demonstrate above average teaching skills? (p. 24).

For some other issues that have been raised about the use of normative comparison groups, see the interchange between Cashin (1992, 1994) and Hativa (1993) and Abrami (1993). Also see McKeachie (1996), who argues that the use of norms can be detrimental to motivation of teachers and are frequently misused in personnel decisions, and a "reaction" by Cashin (1997).

STUDENT RATERS AS INDEPENDENT REPLICATES; COMPOSITION OF STUDENTS IN THE CLASS

When reading a report by Miller, Wilkes, and Cheetham (1993), I was reminded of what I now introduce as a second set of issues. This report describes the experience-based fine tuning of an innovative introductory sequence in first-year biology at Worcester Polytechnic Institute. Although the instructors changed the course from year to year—partially in response to student evaluations of the teacher and course—the next group of students did not necessarily like the course better, and sometimes liked it less (at least on average). In tracking down why these results came about, the researchers discovered that the cognitive style distribution of the student population shifted substantially from year to year, so that any curriculum change that was introduced in response to the previous year's feedback from students was not necessarily appropriate for the current year's class. The researchers write that "Responding to course evaluations is like trying to hit a moving target: The students that one intends to reach may not be there by the time one's aim is adjusted" (p. 44).

In essence, the results reported in this research by Miller, Wilkes, and Cheetham (1993) raise some of the issues I wrote about roughly 20 years ago (Feldman, 1977). I want to resuscitate this earlier article by reiterating, updating, and extending some of the issues it considered. These issues are still pertinent today and will be tomorrow (a case, it seems, of "back to the future"). Put otherwise, the issues have not disappeared since 1977; and, as of yet, they have not been satisfactorily resolved.

I begin by noting that one part of the thinking in the field of student ratings assumes that the traditional model of psychological testing is applicable to student ratings of teachers. In a compellingly argued paper, Crichton and Doyle (1975) not only pointed out this assumption but also analyzed many of its implications. As these analysts note, the traditional testing model deals with a data matrix composed of persons being tested and test items; the items dimension is collapsible into a total-scores vector by assuming that items are replicates except for random error. When adapted or generalized to the rating of instructors, raters (students) take the place of test items and those being rated (instructors) take the place of those being tested (in the assumptions and deductions of the testing model). Thus, it is assumed that students are independent replicates of one another, except for random error, in their "measuring" of the attributes of teachers.

One rationale given for using class average of student ratings, in fact, is that taking the mean of individual observations tends to reduce errors—that is to say, individual idiosyncrasies and limitations, as well as other nonsystematic errors of observation, tend to cancel out. In this respect, as noted, raters are considered as functioning very much as do "items" on conventional tests. Indeed, interrater reliability is related to the number of raters as given by Spearman-Brown Prophecy Formula for test length; increasing the number of raters is viewed as a special type of lengthening. To repeat, student raters are assumed to be independent replicates. This assumption raises a whole series of issues, revolving around whether student raters are indeed *replicates* of each other and whether they are indeed *independent* of each other—and whether they need to be either of these.

Are Students (as Raters) Replicates of One Another?

Whether student raters can be seen as (or indeed are) replicates depends in part on the *stance* they choose to take—or in some cases are more or less required to take—when rating their teachers. By stance, I refer to whether students are "objective" or "subjective" in their rating of teachers. This basic distinction was hardly new even in 1977. In 1973, Sockloff (1973) had already distinguished rating scales from attitude scales. In the same year, Menges (1973) made the distinction between students as reporters and students as judges; and McKeachie (1973) distinguished between asking for students' descriptions of teachers' behaviors and their "assessment of teaching effectiveness or value of the course in their education" (p. 213). Some years after, McKeachie (1979) spoke of items on rating forms that ask students to report their observations of an instructor's behaviors as distinguished from items that are "evaluative."

The Objectivity/Subjectivity of Student Ratings

The degree of objectivity (or subjectivity) of student ratings is a complex matter. At least three dimensions are involved when considering the degree to which student ratings of teachers or courses are objective rather than subjective (or vice versa). First, the items on a rating form or evaluation questionnaire as well as the responses of students can be distinguished by whether students in essence are *neutrally describing* the attributes of teachers and courses or whether they are *giving their evaluative reactions* to them—to name the two ends of a continuum. With respect to the way in which questionnaire items are framed, the distinction is exemplified by the following contrasting items: "To what degree was the course material organized by your instructor?" (to which the student, say, checks one of the following four alternatives: highly organized, somewhat organized, somewhat disorganized, highly disorganized) versus "How satisfied were you with your instructor's organization of the course?" (highly satisfied, somewhat satisfied, somewhat dissatisfied, highly dissatisfied).

Items on a teaching rating form may request a judgmental reaction without explicitly mentioning evaluation. Thus, if the student is asked whether the amount of assigned material for the course was "excessive," "just right," or "too little," the very categories available for response imply the student's judgmental reaction rather than neutral description of the amount of reading required in the course. Moreover, the mere use of the word "describe" or its equivalent does not automatically make the rating nonevaluative, especially if the item's content is global in nature. Thus the student may be asked to "describe" the degree of the teacher's overall effectiveness by marking an appropriate category ("highly effective," "somewhat effective," etc.), but it is unlikely that a "pure" *description* of the teacher's effectiveness will result. Rather, it is more likely that, in part, the student's overall evaluation and degree of *approval* of the teacher and course will be elicited.

The second dimension is the *point of view* that is taken by the student in rating the teacher or course. The distinction here is between taking the point of view of the group of students as a whole (in effect putting oneself into the place of the other students, or the typical student, in the class) or taking a personal view. This second dimension is applicable to either the description or evaluation of the teacher and course (the first dimension). In terms of descriptions, the two extremes of this dimension are illustrated by the following two "stems" of rating items: "Describe the degree to which the instructor stimulated the interest of the class in the course material" versus "Describe the degree to which the instructor stimulated your interest in the course material." (The following wording would represent an ambiguous case, for it is not clear which of the two views is to be taken: "Describe the degree to which the instructor stimulated interest in the course material.") The two ends of this same dimension for students' evaluations or satisfactions are similar to those just given for their descriptions—for example, "How satisfied was the class with the degree to which the teacher stimulated interest in the course material?" versus "How satisfied were you with the degree to which the instructor stimulated your interest in the course material?".

The third dimension (cross-cutting the other two dimensions) is based on the *amount of inferring* students must do—in either their description or evaluation, whatever the point of view that is taken. This is the dimension that Murray

(1983a, 1983b, 1985, 1991; Erdle and Murray, 1986) has elucidated so well. Although there is not a perfect one-to-one correspondence, the degree of inference is generally lower (a) the more visible to the student are the attributes of the teacher to be assessed and the more direct the information that the student has of these attributes, (b) the more a student is asked to consider behavioral attributes of the instructor rather than predispositional or attitudinal attributes, and (c) the more molecular (less molar) the behavior to be rated, or the more specific (less global) the attitudinal attribute of the teacher to be assessed.

Table 1 gives all three (two-category) dimensions, along with their "cross-hatching" and a sample questionnaire item for each combination of categories. To

TABLE 1: Three (Two-Category) Dimensions of Objectivity/Subjectivity with Sample Questionnaire Item for Each Combination of Categories

High(er) inference rating: instructor's organization of the course materials			
Questionnaire item asks for description		**Questionnaire item asks for evaluative reaction**	
Personal point of view	**Students as a whole**	**Personal point of view**	**Students as a whole**
How would you describe the degree to which the instructor organized the materials of the course?	How would the class describe the degree to which the instructor organized the materials of the course?	How satisfied were you with the instructor's organization of the materials of the course?	How satisfied was the class with the instructor's organization of the materials of the course?
Low(er) inference rating: instructor's use of headings and subheadings to organize lectures			
Questionnaire item asks for description		**Questionnaire items asks for evaluative reaction**	
Personal point of view	**Students as a whole**	**Personal point of view**	**Students as a whole**
How would you describe the degree to which the instructor used headings and subheadings to organize lectures?	How would the class describe the degree to which the instructor used headings and subheadings to organize lectures?	How satisfied were you with the instructor's use of headings and subheadings to organize lectures?	How satisfied was the class with the instructor's use of headings and subheadings to organize lectures?

take together the three dimensions being considered, it may be said that the more that evaluation is elicited by the rating item, the more that a personal view is evoked, and the greater is the degree of inference required on the student's part, the less "objective" is the rating. In Table 1, the questionnaire item that theoretically represents the least objective end of the continuum is "How satisfied were you with the instructor's organization of the materials of the course?" The attempt to make ratings more objective is done by asking students to rate descriptively various delimited and visible behavioral attributes of teachers, requiring low inference and a nonpersonal view from them. In Table 1, the questionnaire item representing the combination of these three dimensions is "How would the class describe the degree to which the instructor used headings and subheadings to organize lectures?"

I do not mean to imply by the use of the terms "objective" and "subjective" ratings that one kind of rating is *automatically* better than the other or that one is *necessarily* preferable to the other. I would assume that the closer student ratings are to being "objective" along the three dimensions the greater the interrater agreement (interrater reliability). Yet there is less evidence (one way or the other) about whether this is so than I would have expected; and there is even evidence in one recent study (Benz and Blatt, 1996) that certain of the higher-inference attitudinal rating items can produce more consensus among students than some lower-inference behavioral rating items. Even if it does turn out that objective ratings generally do show higher interrater agreement than subjective ratings, this fact would not necessarily mean that the more objective ratings are automatically more accurate and more *valid* indicators of teaching effectiveness. For instance, Murray (1983b, 1991) has shown that low-inference items are indeed valid in that they are associated with a variety of indicators of teaching effectiveness, but these data do not by themselves show that the low-inference items are more highly related to these indicators than (parallel) higher-inference items.

What Stance Do Students Take?

Students may or may not take the stance either explicitly directed or more subtly implied in the evaluation form. We do not, in fact, know very much about what does go on in students' minds when they fill out rating forms. From existing research we do know something about students' attitudes toward participation in instructor/course evaluation, what students think of the rating forms they fill out, how seriously they take the evaluations, their self-perceived qualifications to rate instructors, their own experiences with faculty evaluation procedures, the factors they think affect ratings, and the like (see, for example, Ingram, 1979; Jacobs, 1987; Purohit and Magoon, 1974; Smith and Carney, 1990; Spencer, 1994; Spencer and Schmelkin, 1994, 1995; Taylor and Ricketts, 1982; Wulff et al., 1985). But we know very little, if anything, about the degree to which the stance taken by students in any particular class is objective rather than subjective (or vice versa)—and how close this stance is to that directed explicitly or implicitly by the rating form itself.

Talley and Timmer (1992) have begun to get one kind of information that would be useful here, although their research is informal in nature. Students in a course in qualitative research methods were given the assignment to talk to their peers regarding the meanings they applied to the questions on an evaluation instrument used at the college. The students were instructed to ask their peers such questions as the following: What do the phrases mean? How are they applied? How do you determine your response? What are you thinking when you evaluate your course and instructors?

A more direct and immediate way of getting at the stance of the student rater would be to interview individual students very soon (if not immediately) after they have filled out rating forms, asking them questions that would establish what stance they took when doing the ratings and what factors seemed to influence their ratings. Something of the sort has been done in a qualitative study by Benz and Blatt (1996). Although these researchers did not ask students directly about the stance they took when completing a faculty evaluation instrument, they did gather information about students' reasoning in giving their ratings and about how the students interpreted the items of the evaluation instrument. Data collection took place simultaneously with the regular end-of-term faculty evaluation. In addition to a standard faculty rating form, the researchers distributed a second form asking about why they rated each item as they did. Among their findings, the researchers report that students used a variety of evidence in making their ratings, attributed their ratings to diverse sources, revealed some interesting understandings about teaching, and felt ambiguous about certain items presented as concrete and quantifiable.

In determining the stance that students take in completing rating forms, it might be of some worth to use a relatively new method from psychology, referred to as verbal report methodology or "protocol analysis" (Crutcher, 1994; Ericson and Simon, 1993; Payne, 1994); in this case, each student (separated from one another) would be asked to think aloud as he or she filled out evaluation forms. Sudman, Bradburn, and Schwarz (1996), in applying methods of concurrent and retrospective protocol analysis to survey methodology, describe how such "thinkalouds" or "talkalouds" give insight into ways in which respondents understand questions, search their memories for relevant information, form judgments, and edit their answers.

Implications of the Stance Taken by Students

The implications of knowing the degree of objectivity (or subjectivity) of ratings are important. If, on the one hand, ratings are meant to be or claimed to be *objective*, then ideally *none* of the differences in the background and attributes of students *within* a particular class nor any of the differences in their experiences in that class should be related to their ratings. Any that do are illegitimate influences; they "bias" results (in at least one of the senses of the term).[4] Clearly students' anticipated grades in the course, the interest and motivation brought by students to the course, and certain attitudinal and personality dispositions of students (or the like) should not be related to objective ratings of the teacher or course. But neither, for that matter, should motivation, interest, and any learning that has actually been induced by the teacher. Even if some students in a particular class, say, were more inspired by the teacher and learned more from him or her (or thought they did), these experiences should not affect neutral and non-personal descriptions and assessments of the teacher's degree of preparation and organization of the course, knowledge of the subject matter, or any of the specific areas in which students are asked to rate their teachers. Indeed, it is arguably the case that these differences in class experiences also should not be related to overall ratings of the general "effectiveness" of the particular teacher, if these global ratings are meant to be neutral, nonpersonal descriptions.[5]

If, on the other hand, ratings are the *subjective* assessments of the teacher—either theoretically or in practice—then teacher-inspired motivation and teacher-induced learning within classes would be expected to be associated with students' evaluation of the teacher's overall "effectiveness" as well as some of the more specific areas of the teacher's performance. Considered more generally, still other characteristics and experiences of students within classes might also be expected to correlate with the student's evaluations of the teacher, given general social psychological theory and research on personal factors that affect individual's perceptions and evaluation of the qualities and behaviors of others.

Researchers who work to improve rating forms and students' responses to them—by creating low-inference items, constructing behaviorally anchored scales, using forced-choice items that are equated on "social desirability," using special rating formats, training students as raters, or the like—essentially are trying to make the forms more objective and to have students take an objective

---

[4]The earlier section on "bias" was restricted to the group level of analysis (that is, to aggregate ratings of courses and teachers) and to attributes of teachers and courses that might create bias in the ratings of students. The characteristics of an *individual* students that might be said to bias his or her ratings is an overlapping but not identical concern (which I do not systematically examine in this chapter).

[5]I am not considering here another set of aspects of individual students that could cause variability in rating a particular teacher (as suggested by Harry Murray, personal communication, February 6, 1996). Even with objective behavioral ratings, it is possible that different students will show different rating-scale behaviors (e.g., some students may tend to give high frequency-of-occurrence ratings for all behaviors, others may tend to give low ratings for all behaviors). Further, the same behavioral description (e.g., "digresses from the main them of the lecture") can mean different things to different student raters (e.g., just exactly what is the main theme, and how far does one have to depart from it before a digression occurs?). Finally, unless it can be assumed that all students attended either all classes or the same subset of classes, then different students may experience different samples of the teacher's behavior. Some of these sorts of realities, too, have been viewed as "biasing" factors or factors producing "errors" in ratings (see, for example, Aiken, 1996; Guilford, 1959, Chap. 7; Thorndike et al. (1991, Chap. 9), but they are of a different type than those being considered here.

stance.[6] And they may come close to the mark (although whether this is so has not yet been fully determined). But the garden variety of rating forms and evaluation questionnaires now in use, and students' responses to them, most probably are not completely objective in terms being discussed here; some subjective components are surely involved.[7]

If some degree of subjectivity can be expected—perhaps more so on some rating forms than on others, or more so on some items than on others on a particular rating form—then some degree of inconsistency among students in their evaluation of teacher is hardly unexpected and might be considered reasonable. Such inconsistency reflects a genuine source of individual differences among students, under the assumption that a given teacher differentially appeals to different students in class. Differences in certain of the attributes and experiences of students may indeed be a source of variation in their appreciation and evaluation of various aspects of the course and teachers, but they are seen as "legitimate inputs to the evaluation process" (Crittenden and Norr, 1973, p. 144) rather than sources of error or "bias" (although we may feel that some differences among students are more legitimate influences than others).

The conclusion, then, is this: to the extent that ratings of students are subjective, any given set of students in a classroom are unlikely to be replicates (cf. Crichton and Doyle, 1975; Feldman, 1977). Before drawing out some implications of this conclusion, I note that the application of the traditional model of psychological testing undoubtedly allows small departures from its assumption of raters as replicates; as these departures increase, however, the use of this pyschometric model becomes less appropriate. I want also to note the possibility of other models supplementing or, to some extent, even replacing the psychometric model. A likely candidate is a model based on survey research (implicitly suggested in Baril, Sebastianelli, and Cannon, n.d.), for one use of surveys is to measure subjective phenomena (Turner and Martin, 1984). Comparisons between psychometrics and survey measurement can be found in Biemer and Stokes (1991), Converse (1984), Groves (1991), and Turner and Martin (1984, Chap. 4); for general explications of the approach and methods of survey analysis, see Biemer et al. (1991), Singer and Presser (1989), Tanur (1983), Turner and Martin (1984).

What would be particularly useful would be research that systematically compares the results of using models based on psychometric and survey measure-

---

[6]For work in this area, see Aiken (1985), Bernadin (1977, 1978), Bernadin, Alvares, and Cranny (1976), Bernadin et al. (1976), Bernadin and Pence (1980), Bernadin and Walter (1977), Borman (1979), Cook (1989), Deutsch (1981), Feldman (1977, 1979), French-Lazovik and Gibson (1984), Halstead (1972), Jako and Murphy (1990), Keaveny and McGann (1975, 1977a, 1977b), Kinicki and Bannister (1988), Kinicki et al. (1985), Kingstrom and Bass (1981), Murray (1983a, 1983b, 1985, 1991), Murphy and Constans (1987), Murphy and Pardaffy (1989), Reardon and Waters (1979), Sharon (1970), Sharon and Bartlett (1969), Waters, Reardon and Edwards (1982).

[7]See Ghiselli and Ghiselli (1972) for an interesting contention that, in principle as well as in practice, ratings in general can never reach (or even approximate) complete "objectivity."

ment. One interesting study along these lines (Baril, Sebastianelli, and Cannon, n.d.) contrasted the use of procedures from two different models to classify teachers as to their degree of instructional effectiveness. The researchers used both the standard error of the mean (inferential statistics as used in the social-behavioral sciences, including survey research) and the standard error of measurement (psychometric model) to classify a group of instructors into high, medium, and low categories based on their overall student ratings. Although the two sets of classifications in this instance were quite similar, the procedure of comparison is informative. The researchers, incidentally, say they prefer a model based on inferential statistics because it "enables the use of the within class variability of each course...[and] provides a particularly rich base of procedures and research applications..." (p. 12). More studies comparing the methods of survey research with those of the psychometric model in the analysis of student ratings would be welcome.

Variability Among Students in a Class in Their Ratings: A First Look

Crichton and Doyle (1975) offered the following analysis about variability in the ratings made by students in class:

> The psychometric literature...reveals a universal attitude of excluding all rater effects from true variance and therefore concluding that reliability means...relative absence of both random error and rater differences. The results of considering raters the source of at least some true variance...must be explored....The traditional theory of reliability of ratings assumes that there exists a true value on a given trait for the ratee which every rater, if he [or she] is not biased or unmotivated or careless or unobservant, will give the ratee. This ignores the possibility that there may be a different "true" value for each student, for example, because the instructor satisfied his [or her] needs or desires with respect to the function named to a differing degree. This would imply the presence of ideographic true variance, true specific rating components of varying magnitude across raters. There should be inconsistencies among raters [even] if they rate without error (p. 19, pp. 27-28).

These various suggestions do not imply that students rate with errors, random or systematic, but that it is analytically possible—and one hopes empirically so—to separate rater error from true rater variance. Researchers and practitioners would still try to eliminate or reduce both random and systematic error by such procedures as making the rating items clear and easy to respond to, by using the most effective rating format, by giving students the same cognitive set in using the rating scales, by trying to give students a uniform level of motivation to respond as well as they are able, and the like.

Now, even if all or some part of the variability among students in their ratings of a teacher is seen as legitimate (as true rater variance), problems still arise—particularly when averaging across these individual ratings. Good reasons exist for the use of average student ratings, of course, both in terms of the increase in reliabilities that result and of the economies gained from data reduction (for purposes of research analysis as well as administrative decisions).[8] And if any diversity of ratings within classes is indeed due to haphazard fluctuations alone, it makes sense to assume that raters are "replicates." However, to the extent that ratings have a subjective component, within-class variability may be more than random error; there may be patterned differences in ratings linked to different types of students in classes, which makes averaging across students' responses less straightforward.

In comparing long-term stability of student ratings of teachers with the within-classroom consistencies among student raters (interrater reliability), Marsh and Overall (1979) present evidence that can be interpreted as showing systematic variance in individual student ratings within classrooms. This longitudinal study demonstrated that, consistent with previous research, the single-rater reliability was generally in the 0.20s for both end-of-course and retrospective ratings (made several years after the course by the same individuals). However, the median correlation between end-of-class and retrospective ratings, when based on responses by individual students instead of class-average responses, was 0.59. About this set of findings, Marsh (1987) writes:

> The explanation for this apparent paradox is the manner in which systematic unique variance, as opposed to random error variance, is handled in determining the single rater reliability estimated and the stability coefficient. Variance that is systematic, but unique to the response of a particular student, is taken to be error variance in the computation of the single-rater reliability. However, if this systematic variance was stable over the several year period between the end-of-course and retrospective ratings for an individual student, a demanding criterion, then it is taken to be systematic variance rather than error variance in the computation of the stability coefficient. While conceptual differences between internal consistency and stability approaches complicate interpretations, there is clearly an enduring source of systematic variation in responses by individual students that is not captured by internal consistency measures. This also argues that while the process of averaging across the ratings produces a more reliable measure, it also masks much of the systematic variance in individual student ratings, and that there may be systematic differences in ratings linked to specific subgroups of students within a class... (p. 277).

The more that such systematic unique variance in student ratings exists, the less sound is the assumption that students are interchangeable and the less easily interpretable are either the averages of student ratings or the reliabilities of these averages (see Crichton and Doyle, 1975; Feldman, 1977).

Even if there are student characteristics that influence individual class-average responses, they would have little effect on the interpretation of class-average responses so long as these characteristics were evenly distributed across classes (Marsh, 1987). However, if the proportional distribution of types of students does vary across classes, variability in the average differences among these ratings may be due to differences in the proportion of various kinds of students in the class as well as to differences in the teachers and courses. As Derry (1979) has

[8]Several other reasons for the use of class means are given in Yunker (1983).

put it, "if ratings are to serve summative purposes...the accidental composition of a class should not have a pronounced effect on a teacher's median [or mean] global ratings" (p. 82). In fact, it is the "pronounced effect" on student ratings in the innovative introductory sequence in first-year biology reported in Miller, Wilkes, and Cheetham (1991) that prompted me to think once again about whether students should be viewed as—or, indeed, must be—independent replicates.

Assuming that in practice most student ratings are not altogether "objective" (just as they are not completely "subjective"), it would seem reasonable to search for ways "to separate the subjective component (depending to some degree on the rater) from the objective component (depending only on what the ratee does) in an individual rating," as Crichton and Doyle (1975, p. 21) put it. These authors suggest that being able to discriminate among the reactions of subgroups of different kinds of students would be a substantial beginning:

> ...perhaps the most realistic strategy to make composite [e.g., average] ratings—and the individual ratings which compose them—more useful would be to try to (a) minimize systematic and random error and (b) then find subgroups within which all total rater contribution (including the error component) approach a constant, or equivalently, in which the observed ratings approach equality....Perhaps the groups will have distinguishable characteristics which will both be an aid in interpreting their ratings and lead to the development or discovery of an external instrument to identify kinds of raters to aid in the interpretation of ratings gathered in the future. Conversely, perhaps it will be possible to group raters according to some theory of how they will rate in a particular situation, and their ratings within subgroups will be more uniform than ratings within the total group (p. 22).

The possibility and attendant complications of subgroupings or subaggregates of students within a classroom rating the same teacher has hardly gone unnoted by analysts. Scriven (1981), for instance, has written that if "only mean ratings are used, then the important case of the instructors who are tremendously successful with a subgroup of the class, perhaps the best students, is overlooked" (p. 253). (Also see, inter alia, Miller, 1984; Chandler, 1978; Centra, 1978; Doyle, 1981; and Tiberius, 1986). Yet, as Marsh (1987) notes, when pointing out that there may be various subgroups of students within the same class who view teaching effectiveness differently and who may be differently affected by the instruction that they receive, "there has been surprisingly little systematic research to examine this possibility" (p. 277).

There is some research, but not much. At least as early as 1932, one study (Wilson, 1932) found that some teachers at the University of Washington had distinctly bimodal distributions of student responses on some of the rating items. That is, more students checked the upper and lower extremes of the categories of responses than checked the central positions. Although the report noted that "an investigation of such cases showed ordinarily that two quite different types of students were in the class," the ways in which the groups differed were not given.

In earlier studies, Centra and Linn (1973) and Singhal (n.d.) also identified subgroups of students within classes distinguished by their responses across rating items or scales. In both studies, these subgroups were identified by obverse factor analysis. Such an analysis uses the subject correlation matrix, in contrast to the item or variable correlation matrix, in order to identify groups of individuals with similar pattern of responses across items or variables. Singhal (n.d.) speculated that the subgroups he found might result from students within the class having different value patterns and experiences, but he did not explore the nature of these values and experiences. By contrast, Centra and Linn (1973) did investigate whether certain characteristics of students would discriminate among the within-class subgroups of students identified in their research. For each of the three courses they studied, a discriminant analysis was run using the subgroups within each class and five student characteristics (expected grade in the course, cumulative grade-point average, year in school, gender, and whether the course belonged to the student's major). In only one of these three courses were any of the discriminant functions statistically significant. Correlating the student characteristics with this function indicated that student year in school and grade expected in the courses were the most highly associated, followed by cumulative grade-point average. Students in groups that were high on the function compared with other students, tended to be freshmen and sophomores, to expect higher grades in the course, and to have higher cumulative grade-point averages.

I did find in my files of research done since my earlier paper (Feldman, 1977) a more recent study by Weber and Frary (1982) that is relevant to the question at hand. These researchers cluster analyzed student ratings in each of 20 classes in order to establish "profile groups" of students in each class. One of the 20 classes had no such clusters; for the other 19, the number of clusters ranged from two to eight. The researchers did not determine the distinguishing characteristics of students in different clusters, however.[9]

In the past two or three decades, "dramatic changes have taken place in the composition of student bodies in American higher education. The diverse elements of today's student body include age, ethnic background, sexual preference, and ever-increasing numbers of "differently" abled, part-time, international, and commuting students" (Smith, 1989, p. 1). As colleges and classrooms increase in the diversification and heterogeneity of their students (Adams, 1992; Border and Chism, 1992; Brookfield, 1988; Feldman and Paulsen (1994); Smith, 1989; Turner et al., 1996), variability in ratings among students in any given classroom conceivably could become more predominant. Whether, in fact, students do cluster in their ratings of teachers should continue to be an important focus of research.

---

[9]Barring these more elaborate analyses, even simpler analyses of patterns of ratings would be useful. For example, is there a narrow or wide dispersion of "scores" around a teacher's mean global rating—and does the amount of this dispersion vary by teacher, type of course taught, or the like? Is the shape of the pattern of scores closer to unimodal or bimodal (the latter indicating a split of students into higher and lower ratings), or does some other pattern prevail?

Variability Among Students in a Class in Their Ratings: A Second Look

One could ask: what of all the studies done at the individual level of analysis that relate student ratings of teachers to the background, attitudes, personality characteristics, and other attributes of students?[10] Do they not qualify as supplying data relevant to the issue of patterned differences in student ratings within the classroom. Mostly no, I would say—or, at best, not unambiguously so. The problem here is that, with occasional exceptions, the analyses in these studies are usually of data pooled across courses; variability in ratings produced by teacher differences and that produced by student differences are thus confounded (as explained in more detail in Abrami, 1985; Cranton and Smith, 1990; Leventhal, Abrami, and Perry, 1977; Linn, Centra, and Tucker, 1975; Perkins and Abbott, 1982; Tollefson and Wigington, 1986).

To parallel more closely the studies that search for student subgroups in the classroom according to similarities of ratings, the correlations of student characteristics with ratings would be done *separately* for each classroom. One could thus determine if results varied by courses and teacher in order to see if it made sense to average the separate correlations across classrooms or even to pool individual student data across classrooms before calculating correlations (cf. Freedman and Stumpf, 1977). In some cases, it might *not* make sense to do so, for it might be found that certain characteristics of students—say the motivation or interest students brought to the classroom—correlated with their ratings in some classrooms and not in others. If so, averaging results or pooling data would confound student effects with teacher effects. It might even turn out that a particular characteristic was related to ratings in opposite ways in different classrooms: for example, perhaps highly motivated students rate their teachers more highly than do lesser motivated students in only certain courses while actually rating their teachers lower than do other students in other courses. Another strategy of analysis would be to use one of the balanced ANOVA designs for statistical analysis suggested by Hopkins (1982); these models allow for the exploration of interaction effects, questions of generalizability and the "legitimacy" of pooling individual observations.

Likewise, the use of hierarchical linear models and similar multilevel data analysis (Bock, 1989; Bryk and Raudenbush, 1992) might prove particularly fruitful in the study of student ratings.[11] Noting that in educational research multilevel modeling and analyses have been mainly focused on primary and secondary institution and students, Ethington (1997) has stressed the importance of multilevel analysis—in particular hierarchical linear modeling—to the study of colleges and college students. Although Ethington has shown how hierarchical

---

[10]For reviews of these studies see Cashin (1988, 1995), Feldman (1976a, 1977, 1993, 1996a), Marsh (1984, 1987), and Marsh and Dunkin (1992).

[11]For some examples of models that have been used and analyses that have been done, see Lee and Bryk (1989), Mason, Wong and Entwisle (1983), Raudenbush and Bryk (1986), and Raudenbush, Rowa, and Cheong (1993).

linear modeling can be applied to the study of the effects of college on students, it could also be applied to the analysis of students' ratings of teachers as well as other areas.

As of this writing, I have come across only one research investigation that has used hierarchical linear modeling in studying the correlates of student ratings: an analysis by Civian and Brennan (1996) of student course evaluations at Harvard College drawn from three consecutive semesters beginning in the spring of 1992. One part of this research sought out the correlates of a composite measure of evaluations of teachers *and* courses. (The researchers refer somewhat inexactly to this component measure as "course satisfaction.") Using a model that simultaneously estimated effects at three levels of data—student, course, and department—Civian and Brennan (1996) found that the way certain student characteristics and perceptions correlated with student ratings varied by course and department (as well as by other conditions):

> We find a main effect—a negative one—for taking a course as a requirement in the humanities (compared to courses in the social sciences). That is, students in humanities courses like the course less if taking it as a requirement than if they are taking it as an elective. Across all divisions, the dissatisfaction of a student taking a course as a requirement...increase[s] as the proportion of students in the class taking it as a requirement increases. Said differently, the more students in a course who are taking it as a requirement, the more dissatisfied the student who is taking it as a requirement compared to the student who is not (p. 10).

The researchers also found that the effect of student-perceived course difficulty varied by division and other conditions:

> In the social sciences, we...[found] on the average a small positive main effect of difficulty: as a student perceives a course to be harder relative to another student who finds it easier, the more s/he likes it. The reverse effect is present for math/science: on the average, finding a course harder is associated with dissatisfaction. On the average, there is no effect of course difficulty in the humanities. In all divisions, however, the magnitude and direction of the within-class effect of difficulty is predicted by the average rating of difficulty for the entire class. For example, in an "easy" course in the humanities, a student likes it better if s/he finds it harder, and in a "hard" course in the humanities, a student likes it less if s/he finds it harder. For all divisions, an increase in the proportion of concentrators translates to an improvement in satisfaction related to course difficulty. So in a course of average difficulty heavily populated by students majoring in the field, a student who finds a course more difficult...[likes] it better than a student who finds it easier. Finally, an increase in the proportion of students taking a course as a requirement translates to a positive effect of difficulty on satisfaction. So, in Core courses that have a high proportion of students taking it as a requirement, finding a course more difficult increases a student's satisfaction relative to someone who finds it easier" (Civian and Brennan, 1996, pp. 10-11).

Along with trying new methods for determining the correlates of student ratings, researchers may also want to explore the possible effects on student ratings of certain student characteristics that have not generally been brought into the analysis of these ratings. The characteristics and attributes that most often have been considered in searching for correlates of student ratings are the student's gender, year in school (college-class level), grade-point average, expected or actual grade in the course, prior interest in subject, and reason for taking the course (for reviews of pertinent studies, see references given in footnote 10). Seldom considered, but possibly of some importance, are various cognitive and motivational characteristics of students that increasingly have been found to be important to their learning and which thus might affect their ratings of teachers. I refer here to cognitive styles, learning styles, and strategies, approaches to studying, perceived (and actual) personal control over important events and outcomes, desirable and undesirable attributions for success and failure, and the like (see Entwistle, 1981; Entwistle and Ramsden, 1983; Entwistle and Tait, 1994, 1995; Fincher, 1985; McKeachie et al. 1990; Menec and Perry, 1995; Perry, 1991; Perry, Menec, and Struthers, 1996).

Are Students (as Raters) Independent of Each Other?

I began this set of issues by asking whether students are independent replicates in their ratings of teachers. So far the argument has been that under certain conditions—perhaps not uncommon ones—they may *not* be *replicates*. That they are not so under these conditions is perfectly legitimate, but it does complicate the study and use of student ratings. Now, to make matters more complex, students are *not* necessarily *independent* raters either (cf. Ronan and Schwartz, 1974). For ratings to be independent, raters should reach their decisions individually rather than as a result of comparing ratings with one another, talking to one another about the ratee, or engaging in formal or informal group discussions and conferences. Otherwise, ratings may become dependent on the personality interaction among raters, the possible influence of the more persuasive or dominant raters, and other such factors that are involved in joint decisions.

In one sense, students do rate their teachers independently. Presumably "conferencing" procedures are not allowed during the actual completion of teacher rating forms; students are usually asked not to compare ratings or to confer with one another. In another sense, however, ratings by students are not altogether independent in typical classroom settings, and may be far from independent in some of them. To one extent or another, students in a class confer throughout the semester about their teacher and the course. They note each other's reactions to the teacher and course material, talk to one another about the teacher and the course, construct with one another the meaning of the teacher's behaviors and interpretations about the teacher's "effectiveness," mutually establish their own "hearsay" about the teacher, and the like. The more difficult and subtle issue, then, is not that of direct collaboration among raters at the time of rating, but rather of mutual influence before the ratings are made and of indirect "contamination" by what is sometimes referred to as the "local reputation" of the ratee, in this case a "reputation" specific to a particular classroom.

The use of the term "contamination" is probably correct if ratings are meant to be objective and one wants to apply the traditional testing model to student ratings. But the term is less appropriate, if not inappropriate, if student ratings are considered as having a legitimate subjective component and if the traditional testing model cannot be applied to these ratings. It can be argued that the immutable reality of the classroom is that students in one degree or another do come to "joint" decisions about the instructor, and this reality should be taken into consideration when analyzing student ratings.

We need to know more than we do about how students arrive at their opinions about any particular teacher and how much (and just how) they have been influenced by their peers. Obviously, much can be learned from interviewing individual students in conventional ways about the matter. For example, Helling, Helling, and Richardson (1978) conducted structured interviews with a 5% sample of seniors who had completed all requirements for graduation at St. Olaf College. Noting that in some classrooms there may be an "organized group definition of the situation," these researchers go on to report:

> Teacher rating scales assume that each responding student is making up his [or her] own mind about the class fairly independently. This turns out to be the exception rather than the rule. In several classes remembered by students [when being interviewed] as the worst classes, activist students have come to a common definition of the situation and have carried on organized protest to the dean or department chairman.... Evidently knowledge of such an organized student protest firmly fixes the definition of a course as a bad one whatever the style of teaching....Such protest does require the presence of a certain kind of student leader in class and if such a person happens not to be enrolled the protest does not happen whatever the provocation. Organized protest is relatively uncommon in the school as a whole but very important in fixing the definition of bad teaching in the minds of students" (p. 7).

Data collection need not be restricted to interviewing students about peer influence on their opinions of their teachers. A number of other methods could be used fruitfully—including using methods of stimulated recall, collecting longitudinal data by administering appropriate questionnaires several times during the semester, having students keep logs or journals during the semester (perhaps combined with end-of semester interviews), "shadowing" selected students through their classes, using focus groups of students from a class and other sorts of group interviews (Noland and Badiali, 1984; Ory, Braskamp and Pieper, 1980; Rippey, 1983; and Wulff et al., 1985).

The methods of data collection just suggested do not necessarily capture *group* processes involved. We need to know more about the group dynamics among students in the classroom to understand better the influence students exert on one another as they form their views of a teacher. Experiments in the social laboratory can supply some useful information—at least one such study (by Bean, 1978) of peer influence on ratings has been done—but study in the classroom itself is needed. Participant observation and a variety of ethnographic procedures and techniques would be useful. Although there is some research on the interactional dynamics in the college classroom, in general it focuses on the interaction between the teacher and students and not on the interaction among students themselves (see, for example, Canada and Pringle, 1995; Constantinople, Cornelius, and Gray, 1988; Ellner and Barnes, 1983; Howard, Short, and Clark, 1996; Karp and Yoels, 1976; Nunn, 1996; Statham, Richardson, and Cook, 1991; Sternglanz and Lyberger-Ficek, 1977).

It might be thought that the dynamics of interpersonal influence among students in the classroom inevitably work toward *increasing* the *consistency* in their ratings of teachers (over and above what it might otherwise be if the students were totally independent raters). Indeed, I wrote something to this effect in my earlier paper (Feldman, 1977). But under certain conditions, interpersonal dynamics may work against increased consistency of student raters. For example, if there are two or more different subgroups of interacting students in the classroom and if each of these groups promotes its own view of the teacher among its student members—with each group thus coming to a different consensus about the teacher—then *less consistency* among ratings would result for the class as a whole. Likewise, the *accuracy* or *validity* of student ratings as a whole (in a classroom) may be greater or lesser depending on the personalities and knowledge of the students in the class, how well students know one another, the communication network among students within and outside of class, whether or not there are informal class "leaders" who help to interpret the meaning of classroom activities and to define the teacher's "effectiveness," and the like. Under certain conditions, then, the interpersonal dynamics among students in a classroom, while making less applicable the traditional psychometric model to student ratings, might well increase the accuracy or validity of these ratings beyond what they would be if student were truly independent raters. Exactly what these conditions are remains to be determined empirically.

CONCLUDING COMMENTS

I began these reflections with remarks on what can be seen as a "continuing quest" in all social and behavioral sciences—namely, establishing the conditions and contexts under which relationships are manifested, are stronger or weaker, and are reversed in direction or otherwise different. Illustrations were given in three areas of interest in the study of effective teaching and student ratings: the connection (if any) between research productivity and instructional effectiveness; the differential influence of specific instructional dimensions on learning outcomes; and the association (if any) between the teacher's gender and instructional effectiveness.

In discussing these three substantive areas of interest, I *assumed* that student ratings are valid indicators of instructional effectiveness. By contrast, when I turned to an analysis of two sets of issues in the field—possible bias in student ratings and the extent to which students are independent replicates in their ratings of teachers and the instruction they receive—I *explored* rather than assumed the validity of student ratings. For the first set of issues, I raised such questions as to how bias can be defined; and when and how to control for (or otherwise take account of) bias so as not to eliminate or ignore legitimate effects of the teacher and course. With respect to the second set of issues, I asked how to separate the objective from the subjective aspects of student ratings; when (and how) to control or adjust for student subjectivity in order to make comparison of ratings across faculty and courses meaningful and fair; and how best to interpret mean scores on evaluation items and multi-item scales (and just when it makes sense to average scores of individual students in the first place). These questions are not easy to answer, and will take more research and thought to arrive at fully satisfactory solutions.

My discussion of certain issues surrounding student ratings has not been done in the spirit characterized by Marsh (1984, 1987) as a "witch hunt." I believe that the use of student ratings is a reasonably sound way of evaluating teachers and courses (although it should not be the only way). I agree with Marsh (1987) that "the reported results [of research] clearly demonstrate that a considerable amount of useful information can be obtained from student ratings; useful for feedback to faculty, useful for personnel decisions, useful to students in the selection of courses, and useful for the study of teaching" (p. 369). As Marsh (1987) points out, students' evaluations of teaching effectiveness are probably "the most thoroughly studied of all forms of personnel evaluations, and one of the best in terms of being supported by empirical research" (p. 369), although he does note that:

> Despite the generally supportive research findings, student ratings should be used cautiously, and there should be other forms of systematic input about teaching effectiveness, particularly when they are used for tenure/promotion decisions. However, while there is good evidence to support the use of students' evaluations as one indicator of effective teaching, there are few other indicators of teaching effectiveness whose use is systematically supported by research findings. Based upon the research...other alternatives which may be valid include the ratings of previous students and instructor self-evaluations [as well as colleague ratings not based on actual classroom observation, as Marsh and Dunkin (1992) further suggest], but each of these has problems of its own (p. 369).

As I explored various measurement, psychometric, and statistical problems of studying effective college teaching and using student ratings, questions of *substance* emerged: What characteristics of teachers, classes, and courses actually affect teaching (whether or not they affect student ratings)? Which sorts of teachers get assigned to teach which sorts of classes? What do students have in their minds when they view and evaluate their teachers? How does the particular composition of students in a class actually affect the instructional effectiveness of teachers (as well as the ratings made by students in the class)? To what extent do students influence one another in their judgments about teachers, and just how do they do so? We have some of the empirical information necessary to answer these questions, but not nearly enough. My intent in this chapter has been to encourage the kind of research that will make student ratings even more useful to educationists than they already are, and, at the same time, will expand our knowledge about social cognition and social attribution (within the field of social psychology) and about teaching and learning in the college classroom (within the study of higher education).

REFERENCES

Abrami, P. C. (1985). Dimensions of effective college instruction. *Review of Higher Education* 8(3): 211-228.

Abrami, P. C. (1989). SEEQuing the truth about student ratings of instruction. *Educational Researcher* 18(1): 43-45.

Abrami, P. C. (1993). Using student rating norm groups for summative evaluations. *Instructional Evaluation and Faculty Development* 13(2): 5-9.

Abrami, P. C., Cohen, P. A., and d'Apollonia, S. (1988). Implementation problems in meta-analysis. *Review of Educational Research* 58(2): 151-179.

Abrami, P. C., d'Apollonia, S., and Cohen, P. A. (1990). The validity of student ratings of instruction: What we know and what do not. *Journal of Educational Psychology* 82(2): 219-231.

Abrami, P. C., d'Apollonia, S., and Rosenfield, S. (1996). The dimensionality of student ratings of instruction: What we know and what we do not. In J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*, Vol. 11. New York: Agathon Press.

Adams, M. (ed.) (1992). *Promoting Diversity in College Classrooms: Innovative Responses for the Curriculum, Faculty, and Institutions*. (New Directions for Teaching and Learning, no. 52.) San Francisco: Jossey-Bass.

Aiken, L. R. (1996). *Rating Scales and Checklists: Evaluating Behavior, Personality, and Attitudes*. New York: Wiley.

Aiken, R. A. (1985). Evaluating ratings on bidirectional scales. *Educational and Psychological Measurement* 45(2): 195-202.

Aleamoni, L. M. (1978). Are there differences in perceived teaching effectiveness between males and females in anthropology? Paper presented at the Annual Meeting of the American Anthropological Association. (ERIC Document Reproduction Service No. ED 176 707).

Baril, G. L., Sebastianelli, R., and Cannon, J.T. (n.d.). The interpretation of student ratings of instructors: An application of inferential statistics. Unpublished manuscript, University of Scranton.

Bean, L. H. (1978). Effects of the social circumstances surrounding the rating situation on students' evaluations of faculty. *Teaching of Psychology* 5(4): 200-202.

Ben-David, J. (1977). *Centers of Learning: Britain, France, Germany, United States*. New York: McGraw-Hill.

Benz, C. R., and Blatt, S. J. (1996). Meanings underlying student ratings of faculty. *Review of Higher Education* 19(4): 411-433.

Bernadin, H. J. (1977). Behavioral expectation scales versus summated scales: A fairer comparison. *Journal of Applied Psychology* 62(4): 422-427.

Bernadin, H. J. (1978). Effects of rater training on leniency and halo errors in student ratings of instructors. *Journal of Applied Psychology* 63(3): 301-308.

Bernadin, H. J., Alvares, K. M., and Cranny, C. J. (1976). Recomparison of behavioral expectation scales to summated scales. *Journal of Applied Psychology* 61(5): 564-570.

Bernadin, H. J., LaShells, M. B., Smith, P. C., and Alvares, K. M. (1976). Behavioral expectation scales: Effects of developmental procedures and formats. *Journal of Applied Psychology* 61(1): 75-79.

Bernadin, H. J., and Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. *Journal of Applied Psychology* 65(1): 60-66.

Bernadin, H. J., and Walter, C. S. (1977). Effects of rater training and diary-keeping on psychometric error in rating. *Journal of Applied Psychology* 62(1): 64-69.

Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., and Sudman, S. (eds.) (1991). *Measurement Errors in Surveys*. New York: Wiley.

Biemer, P. P., and Stokes, S. L. (1991). Approaches to the modeling of measurement errors. In P. P. Biemer, R. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman (eds.), *Measurement Errors in Surveys*. New York: Wiley.

Blackburn, R. T., and Lawrence, J. H. (1995). *Faculty at Work: Motivation, Expectations, Satisfaction*. Baltimore: Johns Hopkins University Press.

Bock, R. D. (1989). *Multilevel Analysis of Educational Data*. San Diego: Academic Press.

Border, L. L. B., and Chism, N. V. N. (eds.) (1992). *Teaching for Diversity*. (New Directions for Teaching and Learning, no. 49.) San Francisco: Jossey-Bass.

Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology* 64(4): 410-421.

Braxton, J. M. (ed.) (1996) *Faculty Teaching and Research: Is There a Conflict?* (New Directions for Teaching and Learning, no. 90.) San Francisco: Jossey-Bass.

Braxton, J. M. and Hargens, L. H. (1996). Variation among academic disciplines: Analytical frameworks and research. In J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*, vol. 11. New York: Agathon Press.

Brookfield, S. D. (1988). *Understanding and Facilitating Adult Learning*. San Francisco: Jossey-Bass.

Bryk, A. S., and Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newburg Park, CA: Sage.

Canada, K., and Pringle, R. (1995). The role of gender in college classroom interactions: A social context approach. *Sociology of Education* 68(3): 161-168.

Cashin, W. E. (1988). Student ratings of teaching: A summary of the research. IDEA Paper No. 20. Manhattan, Kan.: Kansas State University, Center for Faculty Evaluation and Development.

Cashin, W. E. (1992). Student ratings: The need for comparative data. *Instructional Evaluation and Faculty Development* 12(2): 1-6.

Cashin, W. E. (1994). Student ratings: Comparative data, norm groups, and non-comparative interpretations: Reply to Hativa and Abrami. *Instructional Evaluation and Faculty Development* 14(1 and 2): 21-26.

Cashin, W. E. (1995). Student ratings of teaching: The research revisited. IDEA Paper No. 32. Manhattan, Kan.: Kansas State University, Center for Faculty Evaluation and Development.

Cashin, W. E. (1997). Should student ratings be interpreted absolutely or relatively? Reaction to McKeachie (1996). *Instructional Evaluation and Faculty Development* 16(2): 14-19.

Cashin, W. E., Downey, R. G., and Sixbury, G. R. (1994). Global and specific ratings of teaching effectiveness and their relation to course objectives: Reply to Marsh (1994). *Journal of Educational Psychology* 86(4): 649-657.

Centra, J. A. (1978). Using student assessments to improve performance and vitality. In W. C. Kirschling (ed.), *Evaluating Faculty Performance and Vitality* (New Directions for Institutional Research, no. 20). San Francisco: Jossey-Bass.

Centra, J. A. (1983). Research productivity and teaching effectiveness. *Research in Higher Education* 18(4): 379-389.

Centra, J. A., and Linn, R. L. (1973). Student points of view in ratings of college instruction. Research Bulletin RB-73-60. Princeton, N.J.: Educational Testing Service.

Chandler, T. A. (1978). The questionable status of student evaluations of teaching. *Teaching of Psychology* 5(3): 150-152.

Civian, J. T., and Brennan, R. T. (1996). Student and course factors predicting satisfaction in undergraduate courses at Harvard University. Paper presented at the annual meeting of the American Educational Research Association.

Cohen, P. A. (1980). A meta-analysis of the relationship between student ratings of instruction and student achievement. Unpublished doctoral dissertation, University of Michigan.

Cohen, P. A. (1981). Student ratings of instruction and student achievement. *Review of Educational Research* 51(3): 281-309.

Cohen, P.A. (1987). A critical analysis and reanalysis of the multisection validity meta-analysis. Paper read at the annual meeting of the American Educational Research Association. (ERIC Document Reproduction Service No. ED 283 876).

Constantinople, A., Cornelius, R., and Gray, J. (1988). The chilly climate: Fact or artifact. *Journal of Higher Education* 59(5): 527-550.

Converse, J. M. (1984). Attitude measurement in psychology and sociology: The early years. In C. F. Turner and E. Martin (eds.), *Surveying Subjective Phenomena*, Vol. 2. New York: Russell Sage Foundation.

Cook. S. S. (1989). Improving the quality of student ratings of instruction: A look at two strategies. *Research in Higher Education* 30(1): 31-44.

Cranton, P., and Smith, R. A. (1990). Reconsidering the unit of analysis: A model of student ratings of instruction. *Journal of Educational Psychology* 82(3): 207-212.

Crichton, L. I., and Doyle, K. O., Jr. (1975). Reliability of ratings. Unpublished manuscript, Measurement Services Center, University of Minnesota.

Crittenden, K. S., and Norr, J. L. (1973). Student values and teacher evaluation: A problem in person perception. *Sociometry* 36(2): 143-151.

Crutcher, R. J. (1994). Telling what we know: The use of verbal report methodologies in psychological Research. *Psychological Science* 5(5): 241-244.

d'Apollonia, S., and Abrami, P. C. (1987). An empirical critique of meta-analysis: The literature on student ratings of instruction. Paper presented at the Annual Meeting of the American Educational Research Association.

d'Apollonia, S., and Abrami, P. C. (1988). The literature on student ratings of instruction: Yet another meta-analysis. Paper presented at the Annual Meeting of the American Educational Research Association.

d'Apollonia, S., and Abrami, P. C. (1996). Variables moderating the validity of student ratings of instruction: A meta-analysis. Paper presented at the Annual Meeting of the American Educational Research Association.

Derry, J. O. (1979). Can students' ratings of instruction serve rival purposes? *Journal of Higher Education* 50(1): 79-88.

Deutsch, F. (1981). Evaluating teaching effectiveness: Catharsis makes a difference. *Journal of Psychology* 107(1): 147-150.

Dilts, D. A. (1980). A statistical interpretation of student evaluation feedback. *Journal of Economic Education* 11(2): 11-15.

Doyle, K. O., Jr. (1981). Validity and perplexity: An incomplete list of disturbing issues in instructional evaluation research. *Instructional Evaluation* 6(1): 23-25.

Ellner, C. L., and Barnes, C. P. (1983). *Studies of College Teaching: Experimental Results, Theoretical Interpretations, and New Perspectives.* Lexington, Mass.: Lexington Books.

Elton, L. (1986). Research and teaching: Symbiosis or conflict. *Higher Education* 15 (3-4): 299-304.

Entwistle, N. J. (1981). *Styles of Learning and Teaching: An Integrated Outline of Educational Psychology for Students, Teachers and Lecturers.* New York: Wiley.

Entwistle, N. J., and Ramsden, P. (1983). *Understanding Student Learning.* London: Croom-Helm.

Entwistle, N. J., and Tait, H. (1994). Approaches to studying and preferences for teaching in higher education: Implications for student ratings. *Instructional Evaluation and Faculty Development* 14(1-2): 2-9.

Entwistle, N. J., and Tait, H. (1995). Approaches to studying and perceptions of the learning environment across disciplines. In N. Hativa and M. Marincovich (eds.), *Disciplinary Differences in Teaching and Learning Implications for Practice.* (New Directions for Teaching and Learning, no. 64.) San Francisco: Jossey-Bass.

Erdle, S., and Murray, H. G. (1986). Interfaculty differences in classroom teaching behaviors and their relationship to student instructional ratings. *Research in Higher Education* 24(2): 115-127.

Ericson, K. A., and Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data* (Rev. Ed.). Cambridge, Mass.: MIT Press.

Ethington, C. A. (1997). A hierarchical linear modeling approach to studying college effects. In J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research,* Vol. 12. New York: Agathon Press.

Feldman, K. A. (1973). Student assessment of teaching: A selected bibliography. *Improving College and University Teaching* 31(1): 62-63.

Feldman, K. A. (1976a). Grades and college students' evaluations of their courses and teachers. *Research in Higher Education* 4(1): 69-11.

Feldman, K. A. (1976b). The superior college teacher from the students' view. *Research in Higher Education* 5(3): 243-288.

Feldman, K. A. (1977). Consistency and variability among college students in rating their teachers: A review and analysis. *Research in Higher Education* 6(3): 223-274.

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers—what we know and what we don't. *Research in Higher Education* 9(3): 199-242.

Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. *Research in Higher Education* 10(2): 149-172.

Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education* 18(1): 3-124.

Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. *Research in Higher Education* 21(1): 45-116.

Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review and synthesis. *Research in Higher Education* 24(2): 139-213.

Feldman, K. A. (1987). Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: A review and exploration. *Research in Higher Education* 26(3): 227-298.

Feldman, K. A. (1988). Effective college teaching from the students' and faculty's view: Matched or mismatched priorities? *Research in Higher Education* 28(4): 291-344.

Feldman, K. A. (1989a). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators and external (neutral) observers. *Research in Higher Education* 30(2): 137-194.

Feldman, K. A. (1989b). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education* 30(6): 583-645.

Feldman, K. A. (1990a). An afterword for "The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies." *Research in Higher Education* 31(4): 315-318.

Feldman, K. A. (1990b). Instructional evaluation (interview given to Maryellen Weimer). *The Teaching Professor* 4(10): 5-7.

Feldman, K. A. (1992). College students' views of male and female college teachers. Part I—evidence from the social laboratory and experiments. *Research in Higher Education* 32(3): 317-375.

Feldman, K. A. (1993). College students' views of male and female college teachers. Part II—evidence from students' evaluations of their classroom teachers. *Research in Higher Education* 34(2): 151-211.

Feldman, K. A. (1994). Identifying exemplary teaching: Evidence from course and teacher evaluations. Paper commissioned by the National Center on Postsecondary Teaching, Learning, and Assessment for presentation at the Second AAHE Conference on Faculty Roles and Rewards.

Feldman, K. A. (1995). Some unresolved issues in studying instructional effectiveness and student ratings. Address (invited by the Special Interest Group: Faculty Evaluation and Development) presented at the Annual Meeting of the American Educational Research Association.

Feldman, K. A. (1996). Identifying exemplary teaching: Using data from course and teacher evaluations. In M. D. Svinicki and R. J. Menges (eds.), *Honoring Exemplary Teaching*. (New Directions for Teaching and Learning, no. 65.) San Francisco: Jossey-Bass.

Feldman, K. A. (1997). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry and J. C. Smart (eds.), *Effective Teaching in Higher Education: Research and Practice*. New York: Agathon Press.

Feldman, K. A., and Newcomb, T. M. (1969). *The Impact of College on Students*. San Francisco: Jossey-Bass.

Feldman, K. A., and Paulsen, M. B. (eds.). (1994). *Teaching and Learning in the College Classroom* (ASHE Reader Series). Needham Heights, Mass.: Ginn Press, 1994.

Fincher, C. (1985). Learning theory and research. In J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*, Vol. 1. New York: Agathon Press.

Franklin, J., and Theall, M. (1993). Student ratings of instruction and sex differences revisited. Paper presented at the Annual Meeting of the American Educational Research Association.

Franklin, J., and Theall, M. (1995). The relationship of disciplinary differences and the value of class preparation time to student ratings of teaching. In N. Hativa and M. Marincovich (eds.), *Disciplinary Differences in Teaching and Learning: Implications for Practice*. (New Directions for Teaching and Learning, no. 64.) San Francisco: Jossey-Bass.

Freedman, R. D., and Stumpf, S. A. (1977). A model to identify contaminants in the rating process (Faculty of Business Administration Working Paper Series). New York: New York University.

French-Lazovik, G., and Gibson, C. L. (1984). Effects of verbally labeled anchor points in the distributional parameters of rating measures. *Applied Psychological Measurement* 8(1): 49-57.

Gage, N. L. (1991, May 24). Use research university to do research on teaching, suggests education professor [Viewpoint]. *Campus Report* (Stanford University), p. 10.

Ghiselli, E. E. and Ghiselli, W. B. (1972). Ratings—*Kundgabe* or *Beschreibung*. *Journal*

of Psychology 80(2): 263-271.
Groves, R. M. (1991). Measurement error across disciplines. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman (eds.), *Measurement Errors in Surveys*. New York: Wiley.
Guilford, J. P. (1959). *Personality*. New York: McGraw-Hill.
Haladyna, T., and Hess, R. K. (1994). The detection of bias in student ratings of instruction. *Research in Higher Education* 35(6): 669-687.
Halstead, J. S. (1972). Students' ratings of college classroom verbal interaction as related to ratings of instructor teaching effectiveness. Unpublished doctoral dissertation, Purdue University.
Hativa, N. (1993). Student ratings: A non-comparative interpretation. *Instructional Evaluation and Faculty Development* 13(2): 1-4.
Hativa, N., and Marincovich, M. (ed.) (1995). *Disciplinary Differences in Teaching and Learning: Implications for Practice*. (New Directions for Teaching and Learning, No. 64.) San Francisco: Jossey-Bass.
Hattie, J., and Marsh, H. W. (1996). The relationship between research and teaching: a meta-analysis. *Review of Educational Research* 66(4): 507-542
Helling, G., Helling, B., and Richardson, J. (1978). Master variables in student teacher evaluation. Paper presented at the First Congress on Education (Toronto).
Hopkins, K. D. (1982). The unit of analysis: Group means versus individual observations. *American Educational Research Journal* 19(1): 5-18.
Howard, J. R., Short, L. B., and Clark, S. M. (1996). Students participation in the mixed-age college classroom. *Teaching Sociology* (24)1: 8-24.
Hoyt, D. P., and Spangler, R. K. (1976). Faculty research involvement and instructional outcomes. *Research in Higher Education* 4(2): 113-122.
Ingram, L. C. (1979). Broken frame and emergent meaning: The case of student evaluation of teaching. *Human Relations* 32(9): 803-818.
Jacobs, C. (1987). University faculty and students' opinions of student ratings. *Indiana Studies in Higher Education*, No. 55. Bloomington, Ind.: Bureau of Evaluation Studies and Testing, Indiana University.
Jako, R. A., and Murphy, K. R. (1990). Distributional ratings, judgment decomposition, and their impact on interrater agreement and rating accuracy. *Journal of Applied Psychology* 75(5): 500-505.
Karp, D. A., and Yoels, W. C. (1976). The college classroom: Some observations on the meanings of student participation. *Sociology and Social Research* 60(4): 421-439.
Keaveny, T. J., and McGann, A. F. (1975). A comparison of behavioral expectation scales and graphic rating scales. *Journal of Applied Psychology* 60(6): 695-703.
Keaveny, T. J., and McGann, A. F. (1977a). An evaluation of alternate rating forms. Faculty Research Paper No. 158 (Department of Business Administration). Laramie, Wyoming: University of Wyoming.
Keaveny, T. J., and McGann, A. F. (1977b). Reducing contamination in students' evaluation of college professors. Faculty Research Paper No. 161 (Department of Business Administration). Laramie, Wyoming: University of Wyoming.
Kinicki, A. J., and Bannister, B. D. (1988). A test of the measurement assumptions underlying behaviorally anchored rating scales. *Educational and Psychological Measurement* 48(1): 17-27.
Kinicki, A. J., Bannister, B. D., Horn, P. W., and Denisi, A. S. (1985). Behaviorally anchored scales vs. summated rating scales: Psychometric properties and susceptibility to rating bias. *Educational and Psychological Measurement* 45(3): 535-549.
Kingstrom, P. O., and Bass, A. R. (1981). A critical analysis of studies comparing behaviorally anchored rating scales (BARS) and other rating formats. *Personnel Psychology* 34(2): 263-289.

Kremer, J. (1991). Identifying faculty types using peer ratings of teaching, research, and service. *Research in Higher Education*, 32(4): 351-361.

Lee, V. E., and Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education* 62(3): 172-192.

Leventhal, L., Abrami, P. C., and Perry, R. P. (1977). Bogus evidence for the validity of student ratings. Paper presented at the Annual Meeting of the American Psychological Association. (ERIC Document Reproduction Service ED 150 510).

Linn, R. L., Centra, J. A., and Tucker, L. (1975). Between, within, and total group factor analyses of student ratings of instruction. *Multivariate Behavioral Research* 10(3): 277-288.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology* 76(5): 707-754.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research* 11(3): 253-388.

Marsh, H. W., and Dunkin, M. J. (1992). Students' evaluations of university teaching: A multidimensional perspective. In J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*, Vol. 8. New York: Agathon Press.

Marsh, H. W., and Overall, J. U. (1979). Long-term stability of students' evaluations: A note on Feldman's "Consistency and variability among college students in rating their teacher and courses." *Research in Higher Education* 10(2): 139-147.

Mason, W. M., Wong, G. Y., and Entwisle, B. (1983). In S. Leinhardt (ed.), *Sociological Methodology 1983-1984*. San Francisco: Jossey-Bass.

McKeachie, W. J. (1973). Correlates of student ratings. In A. L. Sockloff (ed.), *Proceedings of the First Invitational Conference on Faculty Effectiveness as Evaluated by Students*. Philadelphia: Measurement and Research Center, Temple University.

McKeachie, W. J. (1979). Student ratings of faculty: A reprise. *AAUP Bulletin* 65: 384-397.

McKeachie, W. J. (1987). Instructional evaluation: Current issues and possible improvements. *Journal of Higher Education* 58(3): 344-350.

McKeachie, W. J. (1996). Do we need norms of student ratings to evaluate faculty? *Instructional Development and Evaluation* 15(Spring): 14-17.

McKeachie, W. J. (1997). Good teaching makes a difference—and we know what it is. In R. P. Perry and J. C. Smart (eds.), *Effective Teaching in Higher Education: Research and Practice*. New York: Agathon Press.

McKeachie, W. J., Pintrich, P. R., Ling, Y-G., Smith, D. A. F., and Sharma, R. (1990). *Teaching and Learning in the College Classroom: A Review of the Research Literature* (2nd Ed.). Ann Arbor, Mich.: National Center for Research to Improve Postsecondary Teaching and Learning, University of Michigan.

Menec, V. H., and Perry, R. P. (1995). Disciplinary differences in students' perceptions of success: Modifying misperceptions with attributional retraining. In N. Hativa and M. Marincovich (eds.), *Disciplinary Differences in Teaching and Learning: Implications for Practice*. (New Directions for Teaching and Learning, no. 64.) San Francisco: Jossey-Bass.

Menges, R. J. (1973). The new reporters: Students rate instruction. In C. R. Pace (ed.), *Evaluating Learning and Teaching*. (New Directions in Higher Education, no. 4.) San Francisco: Jossey-Bass.

Michalak, S. J., Jr., and Friedrich, R. J. (1981). Research productivity and teaching effectiveness at a small liberal arts college. *Journal of Higher Education* 52(6): 578-597.

Miller, J. E., Wilkes, J., and Cheetham, R. D. (1993). Tradeoffs in student satisfaction: Is the "perfect" course an illusion? *Journal on Excellence in College Teaching* 4: 27-47.

Miller, S. N. (1984). Student rating scales for tenure and promotion. *Improving College and University Teaching* 32(2): 87-90.

Murphy, K. R., and Constans, J. I. (1987). Behavioral anchors as a source of bias in rating. *Journal of Applied Psychology* 72(4): 573-579.

Murphy, K. R., and Pardaffy, V. A. (1989). Bias in behaviorally anchored rating scales: Global or scale-specific? *Journal of Applied Psychology* 74(2): 343-346.

Murray, H. G. (1983a). Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness. *Journal of Educational Psychology* 75(1): 138-149.

Murray, H. G. (1983b). Low-inference classroom teaching behaviors in relation to six measures of college teaching effectiveness. In J. G. Donald (ed.), *Proceedings of the Conference on the Evaluation and Improvement of University Teaching: The Canadian Experience*. Montreal: Centre for Teaching and Learning Services, McGill University.

Murray, H. G. (1985). Classroom teaching behaviors related to college teaching effectiveness. In J. G. Donald and A. M. Sullivan (eds.), *Using Research to Improve Teaching*. San Francisco: Jossey-Bass.

Murray, H. G. (1991). Effective teaching behaviors in the college classroom. In J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*, Vol. 7. New York: Agathon Press.

Noland, J. F., and Badiali, B. J. (1984). Using student journal entries as a mechanism for evaluation of post-secondary curricula. Paper presented at the annual meeting of the American Educational Research Association.

Nunn, C. E. (1996). Discussion in the college classroom: triangulating observational and survey results. *Journal of Higher Education* 67(3): 243-266.

Ory, J. C., Branskamp, L. A., and Pieper, D. M. (1980). Congruency of student evaluative information collected by three methods. *Journal of Educational Psychology* 72(2): 181-185.

Pascarella, E. T., Edison, M., Nora, A., Hagedorn, L. S., and Braxton, J. (1996). Effects of teacher organization/preparation and teacher skill/clarity on general cognitive skills in college. *Journal of College Student Development*, 37(1): 7-19.

Pascarella, E. T., and Terenzini, P. T. (1991). *How College Affects Students: Findings and Insights from Twenty Years of Research*. San Francisco: Jossey-Bass.

Paulsen, M. B., and Feldman, K. A. (1995). *Taking Teaching Seriously: Meeting the Challenge of Instructional Improvement*. (ASHE-ERIC Higher Education Report No. 2.) Washington, D.C.: The George Washington University, Graduate School of Education and Human Development.

Payne, J. W. (1994). Thinking aloud: Insights into information processing. *Psychological Science* 5(5): 241, 245-248.

Perkins, D., and Abbott, R. D. (1982). Validity of student ratings for two affective outcomes of introductory psychology. *Educational and Psychological Measurement* 42(1): 317-323.

Perry, R. P. (1991). Perceived control in college students: Implications for instruction in higher education. In J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*, Vol. 7. New York: Agathon Press.

Perry, R. P., Menec, V. H., and Struthers, C. W. (1996). How research on learning strengthens instruction. In R. J. Menges, M. Weimer and associates, *Teaching on Solid Ground: Using Scholarship to Improve Practice*. San Francisco: Jossey-Bass.

Purohit, A., and Magoon, A. J. (1974). Congruence in attitude of instructors and students towards course evaluations. *College Student Journal* 8(1): 29-36.

Raudenbush, S., and Bryk, A. S. (1986). A hierarchical model for studying school effects. *School of Education* 59(1): 1-17.

Raudenbush, S. W., Rowan, B., and Cheong, Y. F. (1993). Higher order instructional goals in secondary schools: Class, teacher, and school influences. *American Educational*

Research Journal 30(3): 523-553.

Reardon, M., and Waters, L. K. (1979). Leniency and halo in student ratings of college instructors: A comparison of three rating procedures with implications for scale validity. *Educational and Psychological Measurement* 39(1): 159-162.

Rippey, R. M. (1983). Evaluating teaching. *Educational Psychologist* 18(1): 25-32.

Ronan, W. W., and Schwartz, A. P. (1974). Ratings as performance criteria. *International Review of Applied Psychology* 23(2): 71-82.

Rose, L. A. (1975). Adjustment of student ratings of teachers for extrinsic influences. *Journal of Economic Education* 6(2): 129-132.

Scriven, M. (1981). Summative teacher evaluation. In J. Millman (ed.), *Handbook of Teacher Evaluation*. Beverly Hills, Calif.: Sage.

Sharon, A. T. (1970). Eliminating bias from student ratings of college instructors. *Journal of Applied Psychology* 54(3): 278-281.

Sharon, A. T., and Bartlett, C. J. (1969). Effect of instructional conditions in producing leniency on two types of rating scales. *Personnel Psychology* 22(3): 251-263.

Singer, E., and Presser, S. (eds.) (1989). *Survey Research Methods: A Reader.* Chicago: University of Chicago Press.

Singhal, S. (n.d.). Inter-group differences on Course Evaluation Questionnaire. Research Report No. 262. Urbana-Champaign, Ill.: Measurement and Research Division, Office of Instructional Resources, University of Illinois.

Smith, D. G. (1989). *The Challenge of Diversity: Involvement or Alienation in the Academy?* (ASHE-ERIC Higher Education Report No. 5.) Washington, D.C.: School of Education and Human Development, The George Washington University.

Smith, M. C., and Carney, R. N. (1990). Students' perceptions of the teaching evaluation process. Paper presented at the Annual Meeting of the American Educational Research Association.

Sockloff, A. L. (1973). Instruments for student evaluation of faculty: Ideal and actual. In A. L. Sockloff (ed.), *Proceedings of the First Invitational Conference on Faculty Effectiveness as Evaluated by Students*. Philadelphia: Measurement and Research Center, Temple University.

Spencer, K. J. (1994). Student perspectives on course and teacher evaluations. Unpublished doctoral dissertation, Hofstra University.

Spencer, K. J., and Schmelkin, L. P. (1994). College student views of course and teacher evaluations. Paper presented at the annual meeting of the Northeastern Educational Research Association.

Spencer, K. J., and Schmelkin, L. P. (1995). Student perspectives on teaching and its evaluation. Paper presented at the annual meeting of the American Educational Research Association.

Stark, J. S., and Lattuca, L. R. (1997). *Shaping the College Curriculum: Academic Plans in Action.* Needham Heights, MA: Allyn and Bacon.

Statham, A., Richardson, L., and Cook, J. A. (1991). *Gender and University Teaching: A Negotiated Difference*. Albany, N.Y.: State University of New York Press.

Sternglanz, S. H., and Lyberger-Ficek, S. (1977). Sex differences in student-teacher interactions in the college classroom. *Sex Roles* 3(4): 345-352.

Sudman, S., Bradburn, N. M., and Schwarz, N. (1996). *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology.* San Francisco: Jossey-Bass.

Talley, K. D., and Timmer, D. A. (1992). A qualitative methods exercise: Student and faculty interpretations of classroom teaching. *Teaching Sociology* 20(1): 75-79.

Tanur, J. M. (1983). Methods for large-scale surveys and experiments. In S. Leinhardt (ed.), *Sociological Methodology 1983-1984*. San Francisco: Jossey-Bass.

Taylor, J. S., and Ricketts, M. S. (1982). Student attitudes about participation in instructor and course evaluations. *Journal of Instructional Psychology* 9(1): 18-24.

Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., and Hagen, E. P. (1991). *Measurement and Evaluation in Psychology and Education* (5th Ed.). New York: Wiley.

Tiberius, R. G. (1986). Metaphors underlying the improvement of teaching and learning. *British Journal of Educational Technology* 17(2): 144-156.

Tollefson, N., and Wigington, H. (1986). Teacher-generated and student-generated variability in teacher effectiveness ratings. *Instructional Science* 15: 109-120.

Turner, C., Garcia, M., Nora, A., and Rendon, L. I. (eds.) (1996). *Racial and Ethnic Diversity in Higher Education* (ASHE Reader Series). Needham Heights, MA: Simon and Schuster.

Turner, C. F., and Martin, E. (eds.) (1984). *Surveying Subjective Phenomena*, Vols. 1 and 2. New York: Russell Sage Foundation.

Waters, L. K., Reardon, M., and Edwards, J. E. (1982). Multitrait-multimethod analysis of three rating formats. *Perceptual and Motor Skills* 55(3): 927-933.

Weber, L. J., and Frary, R. B. (1982). Profile uniqueness in student ratings of instruction. *Journal of Experimental Education* 51(1): 42-45.

Wigington, H., Tollefson, N., and Rodriguez, E. (1989). Students' ratings of instructors revisited: Interactions among class and instructor variables. *Research in Higher Education* 30(3): 331-344.

Wilson, W. P. (1932). Students rating teachers. *Journal of Higher Education* 3(2): 75-82.

Wulff, D. H., Staton-Spicer, A. Q., Hess, C. W., and Nyquist, J. D. (1985). The student perspective on evaluating teaching effectiveness. *ACA Bulletin* 53: 39-47.

Yunker, J. A. (1983). Validity research on student evaluations of teaching effectiveness: Individual student observations versus class mean observations. *Research in Higher Education* 19(3): 363-379.

Differentiation and Diversity in Higher Education Systems

Jeroen Huisman
Center for Higher Education Policy Studies, University of Twente

INTRODUCTION

"We have in this country a rich array of institutions that serves a variety of needs. We celebrate the diversity, acknowledging that our system of higher education is the envy of the world....[O]ur goal must be continuously to promote both excellence and diversity in higher education." This quote, from a Carnegie Foundation (1987, p. 2) classification of American higher education institutions, voices the—generally accepted—conviction that a certain amount of institutional diversity is a necessary condition for a flexible, adaptive, and responsive higher education system. Maintaining and increasing diversity has therefore been a major issue in American higher education (Birnbaum, 1983).

Nevertheless, diversity is not solely an American issue. In a number of other countries maintaining (Canada, France, Germany, Switzerland) or enlarging (Australia, Denmark, Japan, the Netherlands, Sweden, United Kingdom) diversity is claimed to be desirable (Goedegebuure et al., 1994). The Australian government, for instance, proposed a restructuring of the higher education system in the late 1980s. Objectives included a more diversified and adaptive system, greater efficiency and partial privatization (Department for Employment, Education and Training, 1988). A large-scale amalgamation of institutions was set in motion, leading to the transition from a binary system to the Unified National System (UNS). Another policy initiative at the beginning of the 1990s (Department for Employment, Education and Training, 1991) reiterates the objective of increasing diversity. Two policy instruments—a quality assurance structure and additional funding for specific national priorities—should ensure diversity: the first by making qualitative differences visible, the second by stimulating the development of institutional missions. In Finland, an increase of diversity was also one of the objectives of a set of recent restructuring proposals. Next to diversifying educational provisions, flexibility and international comparability were objectives aimed at by the Ministry of Education (1992). Whereas the Australian government chose to abolish the college sector, the Finnish government thought

J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*

diversity would best be achieved by introducing a non-university sector. In 1991 the government authorized a number of experiments to provide the new kind of education in fields such as commerce and business, forestry and agriculture, and technology. Policy-makers in Austria felt a similar need for a diverse higher education system. The university sector was considered an extreme of systemic, programmatic, and structural homogeneity (Gruber, 1993). Instead of establishing a new sector, the government focuses on the level and type of programs offered. A new act, introduced in 1993, allows universities and colleges to establish *Magister FH* or *Diplom-Ingenieur FH* programs. An autonomous council (*Fachhochschulrat*) is charged with the establishment of the programs.

It can be observed, first, that governments and other actors in higher education think diversity and differentiation to be important issues. Generally, it is argued that the present-day complex society has a broad range of needs which cannot be fulfilled by one type of institution. Differentiated systems are able to cope better with these multifaceted needs. However, the focus is on different forms of diversity, varying from systemic diversity (diversity related to differences between institutional types: the introduction of a non-university sector in Finland, increasing differences between universities in the Australian UNS) to programmatic diversity (diversity related to level, area, mission and length of programs: the new FH programs in Austria). In the higher education literature the concepts of differentiation and diversity (and related terms such as variety, diversification, heterogeneity, and to differentiate) are used interchangeably and often lack definitions. The variety of approaches calls for a framework that clearly demarcates the different conceptualizations of the terms and their interrelatedness.

A second observation relates to the role of different actors in achieving or maintaining diversity. Although state level actors in the United States have tried to regulate diversity in public higher education systems, in most cases government intervention (by funding initiatives and supporting particular institutions) is less obtrusive than in many European countries (characterized by state level regulation and stricter planning and control). This prompts the question, for policy-makers, how increasing or protecting diversity should or could be best achieved and, for researchers, how processes of differentiation or changes in the level of diversity can be explained.

The two observations above structure the contents of this chapter. First, I discuss the variety of conceptualizations of differentiation and diversity, resulting in a set of three interrelated but distinct concepts. Second, I discuss a review of the research literature on diversity and differentiation in higher education. The state of the art will be assessed, especially focusing on measuring diversity, explanatory mechanisms, and strengths and weaknesses of the approaches. The review will show some important shortcomings. Therefore, third, I present a research agenda to gain substantial insight in causes of diversity and factors inhibiting and stimulating diversity.

CONCEPTUALIZATIONS

The terms differentiation and diversity and related terms all have in common a focus on the existence or emergence of differences, but have various meanings and connotations. Since the terms are derived from biological and ecological theories, I will present an overview of these terms and their meaning in the following sections. Thereafter, the usefulness of these concepts in the context of higher education are discussed. Note that looking at the use of the concepts in biology has no other objective than to gain insight. The biological literature is used as an orienting device, not as a strict, paradigmatic guideline.

Differentiation

In developmental processes in biological sciences, the term differentiation refers to the emergence of several parts from a formerly integrated whole. As an example, consider the development of the human body from the point of fertilization of the egg-cell. Division of cells and growth takes place and throughout the developmental process different parts of the human body emerge, each fulfilling its own function as part of the larger whole.

Two important corollaries of this concept of differentiation exist. First, the emerging parts still "need" each other to be meaningful. Although the heart can be distinguished from other parts of the human body, it has no significant meaning without its relationship to the other parts. This makes clear that in differentiation processes the focus is especially on the function of the whole, despite the visibility of the different parts and the functions of those parts in relationship to the whole. The human body still fulfills the same function irrespective of the developmental phase. In this sense differentiation differs in meaning from terms like segregation and division, which denote a loss of a function of the formerly integrated whole; the emerging parts develop their own function, to a large extent unrelated to the function of the whole. The second corollary is that, because the main focus is on the whole, it is often hard to distinguish the parts from each other within the whole. Although it could be argued, at a more abstract level, that in the given example the emerging parts of the body are "present" at the moment of conception (by means of the genetic blueprints), these parts are not (yet) demonstrable. The boundaries between the whole and its environment are clear, but the exact moment in time that the different parts of the whole begin to fulfill their own function is ambiguous. The differences are only clearly visible when the developmental process has reached its last stage.

Diversity

Contrary to differentiation, diversity refers to a static situation. It refers to the characteristics of a community consisting of organisms of different species. In common language it often expresses solely the variety, i.e. the number of species. In ecological terms it also expresses the events of the distribution of the species

(Huston, 1994; Pielou, 1977): the more abundant a community and the more even the organisms are divided across the species, the larger the diversity of that community. Species are defined by their breeding capacities: organisms that belong to different species are not able to bring forth prolific offspring. This definition is applicable to a large part of biological life, with the exception of animals and plants for which no "intercourse" takes place (vegetative reproduction).

Diversification

Diversification refers to processes in which the diversity of a system increases, whether by means of the growth of the number of species or by means of a change of the dispersion of the organisms across the species. Diversity indices (Ludwig and Reynolds, 1988; Patil and Taillie, 1982) can be used to underline the process of diversification numerically. Diversification can be seen as the dynamic counterpart of diversity. With differentiation, diversification has in common the referral to the transition from one state to another. However, the term diversification (like diversity) does not assume that the constituting organisms necessarily form a whole as in the unity of the human body in the example above. The organisms of the different species in the community are the point of departure, not the community as such. The system in which the diversity can be measured often has no theoretical background; it could vary from a specific geographical area to a system only consisting of a certain type of animal (mammals, birds), and ignoring the presence of other animals. With respect to differentiation the unit (the integrated whole) is the starting-point for analysis of developmental processes. Schematically the similarities and differences between the three terms described above can be depicted as follows (Table 1):

Table 1: Three biological concepts on two dimensions

concepts	static/dynamic	unit of research
differentiation	dynamic	integrated whole
diversity	static	organisms of a community
diversification	dynamic	organisms of a community

The Translation from Biology to Higher Education

How useful are these biological concepts to a review of higher education studies on differentiation and diversity (and related terms)? It seems that the biological terms could be used in a meaningful way, although in practice some problems loom. To set the stage, I indicate what the three concepts, ideally, could represent. Note that the two dimensions used to distinguish the three biological concepts are also used in the context of higher education.

The term differentiation signifies a process in which different structures or functions develop from a formerly integrated whole. The term is, for instance, applicable to a department of a university in which research and teaching were inseparably intertwined, but through time became institutionalized within different structures (units or departments). It is also applicable in situations where the number of entities within a certain system increases. For example, Blau (1973) views the increase of numbers of departments, sections, and administrative levels within academic institutions as processes of differentiation. The term diversity can be used when referring to the variety of types of entities (higher education institutions, study programs, disciplinary cultures) within a certain system (the higher education system, a sector of the system, a university) or to a combination of the variety of types and the dispersion of entities across the types. In this context, the term "type" is chosen to be analogous to species. Finally, the term diversification refers to a process in which a system of types of entities changes into a system that is more diverse, e.g. by introducing new institutions or enlarging differences between universities.

The main problems in the actual use of the concepts in the context of higher education (or any other social context) are threefold.

A general problem refers to the use of all concepts. For a valuable use, clear definitions or descriptions of the entities to which diversity and processes of differentiation and diversification are applied is a prerequisite. Many researchers fail to report the actual units of research to which the processes of differentiation or diversification apply.

A set of three interrelated problems relate to the concept of (measuring) diversity. Whereas in biology the use of the concept is reserved for communities of organisms belonging to different species, in the context of higher education the term can be applied to all classifiable entities (institutions, programs). Unfortunately, the species concept based on interbreeding capacities of organisms is not transferable to any social system. Consequently, expedients must be found to define types of classifiable institutions. This may not be a problem in itself, as will be clear in the review of studies below, but the difficulty involved is that different classifications are possible and even meaningful. Connected to this issue is the fact that a biological organism cannot change its "identity" in terms of the species it belongs to, whereas entities in the social world can and indeed do change their identity (universities change their mission, merge or are renamed, etc.). In the context of higher education, it also seems worthwhile to gain insight into processes that lead to entities becoming more or less similar (trends of homogenization or heterogenization). The concept of diversity in ecology hardly takes into account the internal variety within the species, except for genetic variety within populations, or the fact that species might become similar. The latter is not of interest, simply because there are no biological "mechanisms" that make species more similar or different: if, however, such developments take place, there is general consensus that these are purely random processes.

A problem concerning differentiation relates to a literal application to social events. The question is to what extent it is possible to decide whether new structures or functions *originate* from the integrated whole, as is the case of the biological meaning. In social systems, as will be shown by some studies reviewed below, new functions or structures (e.g. a new type of higher education institution) often are "infused" (Rhoades, 1990) in the system surveyed. In these cases it is hardly possible, or relevant, to decide whether new functions or structures emerge from inside or are introduced from outside the system. It is recommended to avoid the biological connotations and conceive of processes of differentiation as the emergence of "something new" (a new type of higher education institution, a new organizational form, a new educational technology) within a certain system.

Distinguishing Differentiation, Diversity, and Diversification

Despite the problems involved, a distinction between differentiation, diversity, and diversification seems tenable. I propose to reserve the term differentiation for processes in which the number of entities of the subject surveyed increases and for processes in which new entities emerge in the system surveyed. The term diversity should be reserved for indicating the variety of types of entities within a system. Dependent on the goals of the researcher, variety of types might refer to the *number* of types or to the number of types and the *dispersion* of entities across these types. To indicate processes in which diversity increases or decreases or in which entities in a system become different or similar, less relevant in ecology, the terms diversification (or heterogenization) and homogenization seem adequate. Table 2 summarizes the meaning of the concepts.

Table 2: Meanings of the concepts differentiation, diversity, and diversification

concept	meaning
process of differentiation	a process indicating an increase of the number of entities a process in which a specific entity emerges (in a larger unit)
diversity	the variety of types the variety of types and dispersion of entities across these types
diversification	an increase in the number of types an increase in the number of types and/or dispersion of entities across these types an increase of differences between entities or types

The distinction is not only tenable but also worthwhile. The different meanings allow for clarifying the connection between the concepts, especially between processes of differentiation on the one hand and diversification or homogenization on the other. When a new entity emerges within a certain aggregate or system (let's say a new university in a higher education system), we can conclude whether the diversity at the system level increased or decreased. When the university is unique, it certainly increases the diversity (heterogenization). When it is a duplication of an already existing type, it decreases the variety (homogenization). Changes in the level of diversity are easily determined in the examples above. When the constituent parts of a system change profoundly (institutions disappear, emerge and/or change identity) diversity indices can be used to decide upon increasing or decreasing diversity. Thus, a process of differentiation does not necessarily lead to an increase of diversity; it depends on the nature of the process. At the same time, diversity does not only change by means of a process of differentiation. Populations (systems) and consequently its level of diversity can also change by constituents that disappear or change identity.

Table 3: A classification of studies on forms of differentiation and diversity

forms	higher education studies
external diversity	classification, typology, comparison of institutions: Baldridge et al., 1977; Birnbaum, 1983; Carnegie Foundation for the Advancement of Teaching, 1976, 1987, 1994; Lysons, 1990a,b; Smart, 1978; Stanley and Reynolds, 1994; Tight, 1988
internal diversity and differentiation	classification, typology, comparison of disciplines: Biglan, 1973a, b; Smart and Elton, 1982; Stoecker, 1993; Whitley, 1984; typology of study programs: van den Bijtel, 1988; organizational differentiation: Blau, 1973
differentiation of roles and functions	differentiation (development, change) of functions, roles, and structures: Aldersley, 1995; Clark, 1978, 1983; Jones, 1996; Maassen and Potman, 1990a,b; Meek, 1991; Parsons and Platt, 1973; Rhoades, 1983, 1990; Riesman, 1956; Skolnik, 1986.

A Classification of Higher Education Studies on Differentiation and Diversity

For a classification of studies, I will follow the distinction between external and internal diversity (proposed by Birnbaum, 1983 and Stadtman, 1980). External diversity, also termed institutional or organizational diversity, relates to differences between higher education organizations. Internal diversity (and differentiation) relates to differences within higher education organizations. I add the category of studies on differentiation of roles and functions. This category of studies does not fit in with the distinction between internal and external diversity, since roles and functions do not always coincide with (parts of) institutions. Scholars focusing on this form of differentiation often look at the higher education system from a macro-perspective, i.e., the system level is implicitly the level of analysis. Of concern are, for instance, the emergence of new types of institutions and heterogenization, or homogenization trends in higher education systems. The studies on internal diversity relate to differentiation of organizational structures, differentiation/diversity of educational programs, and differentiation/diversity of the academic disciplines. Table 3 gives an overview of the studies. Since the focus of this chapter is on macro-level developments in higher education systems and the latter are, certainly from a policy perspective, viewed as most important[1]. The category of internal diversity and differentiation will not be reviewed (but some examples are mentioned). In the following sections research on external diversity and differentiation of roles and functions is highlighted.

EXTERNAL DIVERSITY

Carnegie Foundation

As has been said, American researchers and practitioners have been concerned about institutional (external) diversity for years. The Carnegie Foundation for the Advancement of Teaching developed the leading typology of American higher education. The foundation proposed a classification of higher education institutions on the basis of the level of degree offered, size, Ph.D. production, research funding, and comprehensiveness of mission. These dimensions resulted in six categories in the latest classification (Carnegie Foundation for the Advancement of Teaching, 1994): *research universities*; *doctoral universities*; *master's (comprehensive) colleges and universities*; *baccalaureate (liberal arts) colleges*; *associate of arts colleges* and *specialized institutions*. The first four categories are subdivided into two parts based on the number of students enrolled, the number of doctoral and master's degrees awarded each year, admission restrictions, and the amount of federal support. The final result is that the classification consists of ten categories. Worthy of note is the fact that the Carnegie Foundation redefined and changed some of the categories over the years. The 1994 classification is very different from the previous two (Carnegie Foundation for the Advancement of Teaching, 1976; 1987). Whereas the classification is mainly a research tool used primarily in sampling and reporting data, it has raised discussions on the extent of diversity of the higher education system in the United States as well as the ways to classify higher education institutions (see below).

Birnbaum: Forms of Diversity and Institutional Types

In a survey of literature concerning diversity, based on Stadtman's (1980) work, Birnbaum (1983, pp. 37-56) identifies seven forms of external diversity:

- **systemic** diversity refers to differences in institutional type, size, and control found within a higher education system;
- **structural** diversity refers to institutional differences resulting from historical and legal foundations, or differences in the division of authority within institutions;

[1]For American higher education this qualification must be explained. In the last decade, debates took place about multiculturalism, i.e. how to cope with the diversity of cultures, ethnicity, gender, etc. in higher education. The focus has been on student access and preparation. More recently the contents and structure of the curriculum are also part of the discussion (see for current debates, e.g.: Marcus, 1994, Smith et al., 1994a,b).

- **programmatic** diversity relates to the degree level, degree area, comprehensiveness, mission, and emphasis of programs and services provided by the institutions;
- **procedural** diversity describes differences in teaching, research, and/or services practices;
- **reputation** diversity communicates the perceived differences in status and prestige;
- **constituential** diversity refers to differences in students served and other constituents in the institutions (faculty, administration);
- **values and climate** diversity is associated with differences in social environment and culture.

To assess the change in institutional diversity in the American higher education system between 1960 and 1980, Birnbaum constructed a typology of institutions based on the following variables: control, size, sex of students, program, degree level, and minority enrollment. Based on various combinations of these criteria, Birnbaum found 141 types in 1960 and 138 in 1980 (in a sample of higher education institutions in eight states). He concluded that "[o]n the one hand, it is true that American higher education has been, and still is, extremely diverse....On the other hand, during a period of unprecedented growth in American higher education, the number of different institutional types has not increased." (Birnbaum, 1983, p. 143). He also noted that "[i]t appears that the higher education system has used the vast increase in resources primarily to replicate existing forms...rather than to create new ones." (Birnbaum, 1983, p. 144).[2]

In the empirical part of his investigation, Birnbaum "only" posed the question of whether diversity increased or decreased. He did not investigate the cause(s). An interesting question is how and if diversity relates to the way in which (state) government steers higher education. Birnbaum (1983, pp. 149-182) argues, for instance, for relaxing rigid criteria for approval of new programs and institutions, limiting state planning, and flexibility in governmental procedures. Planning is, according to Birnbaum, a threat to institutional diversity because, first, it restricts experimental innovations of institutions their search for fitness. Second, state-level planning does not reflect knowledge about how institutions adjust to their niches. Third, state-level planning leads to centralization, which paves the way for homogenization of norms, values, and structures and thus decreases diversity. When state government wishes to maintain diversity, governmental actions should be characterized by stimulative and encouraging policies.

Birnbaum's hypotheses have unfortunately not been tested. But a counterexample makes clear that governmental interference does not always threaten institutional diversity. The California system of higher education is an example of a

[2] Zammuto (1984) investigates, in a similar vein, the changes in the population of colleges and universities, and especially those in the subpopulation of liberal arts colleges between 1972 and 1981. The diversity decreased at the population level and remained constant within the subpopulation.

system in which state government, by means of conscious legislative decisions, attempts to maintain a tripartite—diversified—public sector (see Fox, 1994; Organization for Economic Co-operation and Development, 1990), although there are indications of tensions between the different types of institutions leading to homogenization (Goedegebuure et al., 1994, p. 318).

Classifications Based on Effectiveness, Professional Autonomy, and Incentives

The following studies discuss external diversity by means of developing classifications of other dimensions than those used by the Carnegie Foundation for the Advancement of Teaching (1976, 1987, 1994) and Birnbaum (1983).

Lysons (1990a,b) studied organizational effectiveness and the organizational climate of higher education institutions in Australia. Lysons sought—following Cameron (1978, 1981)—a taxonomy to rank higher education organizations based on their effectiveness. Relevant to this study, Lysons' taxonomy is useful in examining if and to what extent higher education organizations form a heterogeneous group. Using factor analysis, Lysons constructed dimensions of the effectiveness of, for instance, the professional development of personnel and the power to attract students of high quality. These dimensions appear to coincide with two other dimensions that were chosen *a priori*: institutional type (universities versus institutes and colleges) and institutional age (old versus young). As a consequence four distinct types of organizations were to be found: older/larger universities, younger/smaller universities, institutes of technology, and colleges. Further evidence for the taxonomy could be found in a study by Goedegebuure et al. (1993). Institutional management's perceptions and attitudes regarding the changes in Australian higher education indeed reflected differences in organizational type, hypothesized by Lysons' classification.

A similar research design can be found in Baldridge et al. (1977), who posed the question whether the 1976 Carnegie Classification did justice to the diversity of higher education institutions in the United States. The researchers developed a new empirical typology based on important characteristics of organizations: the relationship with the environment, features of professional tasks, and the complexity of the organizations. Using these key factors, the Carnegie Classification could be reduced to eight types. The study also indicated a connection between the Baldridge typology and the extent of professional autonomy, defined as the ability of the faculty to control their task environment. The study concludes that the American higher education system is to a large extent diversified using the organizational characteristics mentioned above as discriminating variables.

Whereas Baldridge et al. (1977) take up the point of the professional autonomy, and Lysons (1990a,b) chooses organizational effectiveness, Smart (1978) focuses on academic personnel and the structure of incentive systems for them. Incentives such as a higher salary, a higher position in the organization, and a decrease of the pressure of work play an important role in faculty and administrative behavior. Using data from a study of incentives, Smart (1978) developed a modified typology derived from the 1976 Carnegie Classification: *comprehensive institutions*, *major research institutions*, *restricted-scope research institutions*, *liberal arts colleges*, and *two-year institutions*. In fact, he confirmed the usefulness of the Carnegie Classification for studying faculty incentives.

Other Typologies: Quantitative Indicators

In the previous section, researchers used questionnaire data to construct classifications and taxonomies. Stanley and Reynolds (1994) clustered the Australian universities based on two data sets consisting of evaluative ratings and quantitative performance indicators, respectively. Performance indicators included are, for instance, number of academic staff per discipline, funding sources, number of completions, and enrollments of different types of students. Evaluative ratings relate to, for instance, breadth and depth of course offerings, admission flexibility, graduate salaries, and employment prospects. Cluster analyses performed on the two data sets revealed that there was not much consistency across the sets. The groupings emerging from the cluster analyses also did not reflect the 'natural origins' (original universities, newer universities, institutes of technology, and amalgamated institutions) of the higher education organizations. The authors argue that the results support the view that the UNS maintained and possibly increased diversity, rather than led to uniformity of institutions.

Tight (1988) mainly used data on student body characteristics of institutions (student numbers, broken down by level, mode, and subject of study) to construct typologies of the English university sector, the polytechnic sector, and the college sector separately, as well as all institutions together. Cluster analysis of the institutions of the university sector revealed five distinct groups and two unclassifiable institutions. The analysis confirmed the stability in the university sector, for the results compare with previous studies (Dolton and Makepeace, 1982; King, 1970). A cluster analysis of all English institutions led to ten main institutional groupings, in which the three sectors of the higher education system were recognizable in four, two, and four groupings, respectively. Despite the reflection of natural groupings of higher education organizations, the study also points at similarities between institutions across the groupings: for instance, some colleges are demonstrably similar to campus universities. This leads to the conclusion that the three basic institutional types may not be as distinct as is often assumed.

Conclusions

The latter five studies (Baldridge et al.(1977), Lysons (1990 a, b), Smart (1978), Stanley and Reynolds (1994), Tight (1988)) are similar. Each classification is based on important—but often invisible—organizational characteristics: effectiveness, professional autonomy, incentive systems, and quantitative (performance) indicators. It is important to note that the choices of these characteristics are well-considered, as is the case in Birnbaum's study. The Carnegie Classification seems to be based—more than the other—on administrative criteria (size, number of Ph.D.'s, amount of federal support, etc.).

Each of the studies also uses a similar research methodology, using statistical techniques (such as factor analysis, discriminant analysis, and cluster analysis) applied to empirical data to develop typologies. This empirical orientation distinguishes these studies from research on differentiation of roles and functions discussed in the following section.

In terms of the conceptualizations put forward in this chapter, most studies (especially Smart, 1978 and Baldridge et al., 1977) examine the *level of external diversity* at a specific moment, by looking at the variety of types of higher education institutions and/or the dispersion of institutions across the types. Some of the studies relate to change over time. Whereas Stanley and Reynolds did not research developments over time, they nevertheless conclude that diversity—presumably—did not increase. Although the Carnegie Foundation for the Advancement of Teaching (1976, 1987, 1994) examines classification trends, it is difficult to draw conclusions about decreases or increases of diversity because the classification categories were redefined over time. Two studies can be seen as "real proof" of developments over time. Tight's (1988) study confirmed the stability of the university sector between the late 1960s and the beginning of the 1980s. Birnbaum's (1983) study is the only one that focuses on the increase (or decrease) of institutional diversity applying a comparative and consistent methodology to longitudinal data.

DIFFERENTIATION OF ROLES AND FUNCTIONS

The studies in the previous section all concentrated on institutions of higher education systems. The studies in this section focus on institutions as well as on types of institutions and sectors of the system. A common theme in the studies are the different roles and functions that the (types of) institutions fulfill in the higher education systems and the forces changing these.

Parsons and Platt: The American University

One of the leading theorists of the structural-functional approach is the late Talcott Parsons. In this section I focus on his work on differentiation in higher education. Differentiation is—according to Parsons (1966, p. 22)—a process whereby "[a] unit, sub-system, or category of units or sub-systems having a single, relatively well-defined place in the society divides into units or systems (usually two) which differ in *both* structure and functional significance for the wider system." In this respect his view on differentiation comes close to that of biologists.

Parsons (1978) and Parsons and Platt (1973) distinguish between the differentiation of the university from other sub-systems of the social and cultural system and the differentiation of the university itself. The social and cultural system are subsystems with their own functions of the general—and abstract—*systems of action*. First, I discuss Parsons' view on the emergence of the American university.

The emergence of the higher education system, its function, and its purpose is the main focus of Parsons and Platt's (1973) theoretical work on the American university. Although the title suggests that the authors consider the whole university system, they actually focused only on the "full" universities; a dozen to fifteen elite institutions, consisting of almost all representatives of the disciplinary range. The study is mainly descriptive-analytical, portraying the relationship of the academic system with other *systems of action*. Parsons characterizes the academic system as the institutionalization of the *cognitive complex*: "As an institutional complex, the university holds fiduciary responsibility for the maintenance, transmission, and development of knowledge in particular, and of cognitive functions and resources in general." (Parsons, 1978, p. 139). In the terminology of the system of action, the cognitive complex is the zone of interpenetration of the fiduciary subsystem of the social system (with the rationality system as the main linking subsystem) and the cultural system (especially the subsystem of cognitive symbolization).

The development of education (Parsons and Platt termed this the *Educational Revolution*) eventually led to the expansion of higher education. With the establishment of the university system, "the cognitive system as institutionalized in the academic world has become a differentiated system with substantial autonomy vis-à-vis other subsystems but also interdependent with them in new ways." (Parsons and Platt, 1973, p. 46). Several other forms of differentiation took place during the *Educational Revolution*. For instance, family and educational roles were differentiated by replacing the parents' socializing function with instruction by professionals (teachers) in an institutionalized setting. Another form of differentiation is the secularization of higher education, its separation from religion: "...a continuing decline in legitimacy of the claim of particular religious positions to monopolize legitimate intellectual points of view." (Parsons and Platt, 1973, pp. 276-277).

Next to the processes in which the higher education system separated from other systems, within the higher education system processes of differentiation (and integration) can be recognized. Almost all American universities, with the exception of the recent period, began as undergraduate colleges. The introduction of new functions—e.g. the development of graduate schools at the beginning of the twentieth century—led to internal differentiation of the university. The graduate schools have come to be differentiated from the undergraduate college but coexisted to a large extent within the same structure. The graduate schools concentrated on "pure" as distinguished from "applied" research (or knowledge for its own sake versus knowledge for problem-solving, see top left of Table 4). The primacy of cognitive rationality of the university (the institutionalization of the cognitive complex) is combined with other values, such as the socialization of generally educated citizens in the undergraduate colleges of the university (the utilization of cognitive resources as an expression of the economic value of the universality of education, bottom left of Table 4). Like the graduate schools, professional schools were absorbed by the university. The applied professions combine practical goals with the values of the wider society (bottom right of Table 4). Although the functions of the university became more differentiated, the actual process of the inclusion of the graduate schools in the university and the emergence and growth of the professional schools at the same time indicate a process of integration of functions within the walls of the university (Parsons, 1978, p. 100). In this respect, Parsons uses the term "university bundle" to stress the highly differentiated structure as well as the integrative forces connecting the distinctive functions.

Table 4: The principal functions of the American university
(Parsons and Platt, 1973, p. 92)

	Knowledge "for its own sake"	Knowledge for "problem-solving"
Institutionalization of the cognitive complex	The core of cognitive primacy (research and graduate training by and of "specialists")	Contributions to societal definitions of the situations (by "intellectuals" as "generalists")
Utilization of cognitive resources	General education of "citizenry" (especially undergraduates as "generalists")	Training of professional practitioners (as "specialists")

Although Parsons and Platt's (1973) analysis—despite the generalized approach, abstractness, and doubtful tenability in present-day American higher education—might be applicable to the historical development of elite institutions, it does not account for the changes in the other parts of the higher education system (see e.g.: Brint and Karabel, 1989; Cowley and Williams, 1991; Geiger, 1986). The following studies in this section try to pay attention to all sectors or institutions within national systems of higher education.

Riesman: The Academic Procession

Riesman (1956) pictured the American system of higher education as a kind of reptilian procession—referring to the movements of these animals, in which the body and the tail follow the head and at any given time will be found at the same place where the head was before. According to Riesman, lower status academic institutions try to gain status by imitating high status institutions. Therefore, the tendency is to move away from diversity toward uniformity, toward the standards of excellence of the prestigious research universities. The argument was elaborated in *The Academic Revolution* (Jencks and Riesman, 1968). Jencks and Riesman argued that this revolution (the professoriate's rise to power) was based on increased professionalism and the university college as the basic model for all higher education institutions. As a consequence, institutions moved away from their original mission toward norms of achievement, competence, and judgment, typical for the academic values of national elite institutions. The driving force behind tendencies of homogenization—according to the authors above—seems to be the academic norms and values (see also Neave 1979, 1983, for similar arguments and European examples). Recently, Jencks and Riesman's argument was confirmed by, Aldersley (1995), who analyzed the change in four categories of institutions of the Carnegie Classification (*research universities I* and *II*, *doctoral universities I* and *II*) between 1976 and 1994. Most institutions that changed category from 1976 to 1994 did so in an upward direction: 127 from the 136, excluding institutions that went up and down or *vice versa* in that period. He concludes that the phenomenon of upward drift continued until the 1990s, and to the same degree as in the late 1970s and the beginning of the 1980s. Note that the results relate to changes in the upper range of institutions. What happened at the other end of the range (specialized institutions, baccalaureate colleges, etc.) was not investigated.

Whereas Birnbaum pointed at governmental policies as the main inhibiting factor for increasing diversity, and Riesman and followers stressed the academic norms and values, the following authors include both factors in their research.

Rhoades: Political Competition and Differentiation

Rhoades (1983, 1990) compared the higher education systems of four countries between 1960 and 1980: the United States, England, Sweden, and France. His proposition is that differentiation is the result of political competition and state action, and that homogenization (*dedifferentiation*) is the "natural" trend in higher education. Rhoades investigated differentiation by analyzing the roles of the different actors in these processes: "And only through considering peoples' beliefs and actions can we explain the tendency toward dedifferentiation in higher education. Human agents determine both the pace and path of differentiation." (Rhoades, 1990, p. 189). In this respect Rhoades agrees with the ideas of sociologists like Eisenstadt (1964) and Rueschemeyer (1977), who also call for attention to interest groups and power in processes of differentiation and dedifferentiation.

Before proceeding, let me consider Rhoades' (1990, pp. 191-192) conceptualization of differentiation. He argues that the emergence of new structures and functions is important in processes of differentiation. He refers to Smelser's (1959, p. 2)—in his eyes inadequate—definition of differentiation: "[S]tructural differentiation is a process whereby one social role or organization...differentiates into two or more roles or organizations....The new social units are structurally different from each other, but taken together are functionally equivalent to the original unit." Smelser—like Parsons—excludes the emergence of totally new functions (i.e., functions that were not part of the original unit), from his definition of differentiation. In contrast, according to Rhoades (1983, p. 285) differentiation should be conceived "...in terms of splitting up existing functions or adopting new, distinct roles for higher education and in setting up distinct institutions geared to these."

Rhoades also argues that two differentiated functions need not perform the same function as the original. When differentiated functions become institutionalized over time (for instance, the introduction of the school as the institutional setting for educating the younger generation differentiated education and family functions), they do not necessarily fulfill the same function as before. Rhoades (1990, p. 192) proposes to view differentiation as a process of change in which *infusion* is important: "Change in higher education may involve the 'infusion' of new concerns and the construction or recasting of organizational units to work in accordance with these." In this respect, Rhoades departs from the "older" sociological conceptualizations of differentiation, which stayed close to the biological concepts (see also Durkheim, 1964; but especially Spencer, 1898).

Returning to his research, Rhoades states that a number of developments took place in the higher education systems of the four countries that should have led to an increase of differentiation. In all four countries, economic growth and state support for higher education decreased after a period of growth in student enrollments. As a consequence, competition between institutions increased, as did the opportunities for differentiation. Each federal government contributed to the environment for differentiation by means of introducing new institutions or types of institutions. Examples are the CAT's (Colleges of Advanced Technology, introduced in 1956) and polytechnics (1969 - 1973) in England, the IUT's (*Instituts Universitaires de Technologie*, 1966) in France, and the community colleges (1963-1971) in Sweden. On the other hand, several factors may have inhibited the increase of differentiation. Government, external organizations (e.g. accreditation boards), and higher education institutions themselves have had their share in homogenization processes. Academic professionals played an especially important role: "I found repeated instances of the profession's stiff resistance to a variety of reforms that could have effected differentiation." (Rhoades, 1983, p. 317). Another, rather diffuse, interest group, labeled lay groups or laity, indicating actors that are not professionally involved in higher education (e.g. economic associations, local communities, ethnic groups, students, and trade unions), also seemed to be of influence. Lay groups involved in decision making procedures and implementation of innovations, had a positive effect on the increase of differentiation.

Rhoades concludes his research by stating: "The receptiveness of a higher education system to differentiation will be a negative function of the power differential between the laity and the academic profession—between the academic interest interpretations and emphases and the challenging interpretations and emphases introduced from outside the higher education system." (Rhoades, 1990, p. 212). The academic profession defends with success its own norms and values and achieves dedifferentiation. The input of laity (and the state) is sometimes necessary to stress other interests in the system and to increase differentiation. Especially in England and France the academic profession had a strong influence in promoting dedifferentiation. In the United States however, thanks to the strong representation of laity, a political structure aiming at institutional missions and a diverse (thus less organized) academic profession, there was sufficient support for differentiation processes.

Clark: Differentiation and Interest Groups

Clark (1978) constructed dimensions of differentiation to carry out comparative research in higher education: "Because of the growing complexity of bodies of knowledge and related tasks, the division of academic labor is increasingly characterized by fragmentation within and among universities, colleges, and institutions." (Clark, 1983, p. 70). The two dimensions of interest to Clark are vertical—horizontal differentiation and differentiation within—between institutions. From these dimensions, Clark derives four differentiated units: tiers, hierarchies, sections, and sectors (see Table 5).

Table 5: Dimensions and units of academic differentiation.

	within institutions	**between institutions**
vertical	TIERS: undergraduate, graduate, professional school	HIERARCHIES: status, prestige
horizontal	SECTIONS: faculty, school, college, chair, department	SECTORS: public vs. private, university vs. non-university

Clark emphasized that the study of differentiation must not be limited to describing the partition of people across parts of the organization, but should include the processes of power and power legitimacy: "[T]o study academic differentiation is not only to determine the academic division of labor in its specific operational settings. It is also a pursuit of the expression of academic values and the foundations of academic power." (Clark, 1978, p. 258).

In a cross-national study, Clark (1983, pp. 214-227) connected differentiation processes to processes of change. He argues that the increased complexity of higher education systems is related to the increased complexity of the tasks the system must fulfill. Increased complexity is a function of three interrelated forces: the increase of the variety of the student population, the growth of the labor market for higher education graduates, and the emergence of new disciplines. These forces correlate with the interests of individuals and groups within the units of the higher education institutions: "Much change occurs through differentiation; differentiation is driven in the immediate setting by the rearrangement of interest; interest is basically divided between those already vested and those seeking to become vested; the outcomes of the struggle are determined by relative power; and power is rooted in respective legitimacies." (Clark, 1983, p. 218).

Over time, Clark (1983, p. 221) assumes an increase of differentiation:

"[O]nce created and made valuable to a group, often to an alliance of groups, academic forms persist. Out of successive historical periods come additional forms, with birthrate greatly exceeding the death rate. Differentiation is then an accumulation of historical deposits." In this respect Clark refers to institutions within the system as well as to (sub)units within the institutions. The persistence of forms (the term comes from Stinchcombe, 1965) emerges from the institutionalization of the ideology of the academic organization in a largely non-competitive setting.

Clark also found support in the work of other sociologists. Durkheim views differentiation as a form of group protection. Durkheim explains the division of labor in society by pointing at the function the division has for society at large. When individuals or groups perform a similar task, they (can) chose to differentiate tasks (e.g. specialization). While performing different functions, individuals or groups become mutually dependent, which strengthens the ties between them with positive effects for the group or for society as a whole.

According to Clark, dedifferentiation also can be a means of group protection (which, by the way, negates the explanatory power of group protection in cases of differentiation). Dedifferentiation is especially evident in academic drift, where institutions lower in prestige try to emulate higher status institutions (often the status of the university). Examples can be found in the history of the four year public colleges in the United States and in the technical colleges in England.

The question becomes how institutions try to defend their interests. Institutions have the choice of distinguishing themselves from other institutions or of protecting themselves by imitation behavior: "Self-differentiation can be highly rewarding, but it is hard and even risky work." (Clark, 1983, p. 223). According to Clark (1983, p. 221), the choice for the way in which institutions protect their interests depends on the legitimacy of institutional roles: "Where plural roles have been made acceptable—to student clienteles, external supporters, and primarily the faculty and administration of each segment—legitimacy is an anchor for differentiation, holding it in place and stabilizing new segments as they emerge. When only a single role is acceptable, then legitimacy of form encourages dedifferentiation." Governments and markets play important roles in differentiation processes. Both can supply the context for the legitimacy of distinctive roles. The government can use the tools of regulation and (financial) stimulation to achieve differences between sectors of the system. In systems where government has less influence on higher education, market competition can stimulate institutions to search for and find their own niche.

Meek: Diversification and Homogenization in Australia

Like Clark, Meek (1991) connects processes of change to the concept of differentiation, although he uses the terms diversification and homogenization. Meek hypothesizes that the two processes have a dynamic as well as a symmetric relationship. For the former, a system may be structurally differentiated but have all institutions serving more or less homogeneous educational functions (Meek, 1991: 474). Symmetry applies when periods of homogenization and diversification alternate. No direction—from homogeneous to diversified or the other way around—should be assumed: "Diversification may disadvantage some groups and advantage others, resulting in tensions which may lead to the demise of the diversified structure and its substitution by a more homogeneous one. The homogeneous state of affairs, once installed, may in turn cause other interest groups to mobilize." (Meek, 1991, p. 475). Similar to Clark (1983) and to Goedegebuure and Westerheijden (1991), Meek explains the occurrence of diversification and homogenization by a struggle between interest groups: "[C]hange is based on power relations and the articulation of interests by various groups whose actions and interests are themselves either constrained or furthered by the structure of the academic field and their location in it." (Meek, 1991, p. 463).

Using this framework, Meek analyzes the demise of the binary system and a large scale merger operation in Australian higher education. He argues that the policy initiatives may not necessarily lead to diversification. Diversification could be inhibited by the prevalent norms and values of higher education, where research is valued more than education and universities consequently receive more financial support than the colleges (CAE's). In this context, lower status institutions emulate higher status institutions to reach university status. Contextual factors (competition, institutional autonomy) rather than policy initiatives may influence diversification: either institutions are challenged to choose a distinctive profile or mission, or institutions show imitative behavior (*mimetic isomorphism*; see DiMaggio and Powell, 1983 and below) which leads to homogenization.

Skolnik and Jones: Diversity in the Canadian Higher Education System

Skolnik's (1986) evaluation of the Canadian higher education system found that the non-degree sector is highly diversified, but the university sector is less so. Skolnik took into account the educational process, the administrative and governance arrangement, the degree of specialization, and the relationship between government and institutions. In this respect the non-degree sector displays more variety than the university sector. The lack of variety in the university sector could be—drawing heavily on Birnbaum's (1983) restrictive factors—a result of the uniformity of governmental regulations for program development and change, and the low degree of student mobility. The (provincial) governmental steering approach (e.g. visible in the authority to grant degrees and the discouragement of private higher education) also restricts differentiation. Finally, and most importantly, academic norms and values act to reduce variation across institutions. Skolnik's (1986, p. 31) conclusion: "While Canada does not have strong traditions of diversity in its university system, it is faced with similar pressures toward homogenization as is the American system. The foremost of these may be a professoriate which identifies nationally, or indeed internationally, with their discipline, and which shares a common set of academic and professional values and norms."

Jones (1996) mentions the following reasons to explain the low level of diversity in the university sector. First, like Skolnik (1986), he stresses the public monopoly of provincial governments to grant degrees. A second influence stems from the geographical size and uneven distribution of the Canadian population; governments have chosen to create internally diversified universities that are roughly comparable in terms of structure and function. From a perspective of equality of access and efficiency, this seemed to be a better solution than a set of highly dispersed and diverse institutions. Note that this observation is also a point of criticism on "diversity as an inherent good", a rather pervasive connotation of diversity. The third reason is the fostering of intergovernmental learning and comparisons by the federal structure. Different provincial governments have in the past reacted in remarkably similar ways to emerging problems and changes (Cameron, 1991). Fourth and finally, because of the relatively small size of the university sector and the broadly similar roles of the institutions, the universities have been able to learn from each other by means of national associations and other networks. For instance, successful innovations are easily adopted by others. An interesting point in the context of this chapter, stressed by Jones, is the fact that a rather homogeneous university sector and a highly diversified non-degree sector can both emerge within a system lacking strong centralized authority or national policies.

Maassen and Potman: Institutional Isomorphism in the Netherlands

Maassen and Potman (1990b) investigated changes in Dutch higher education, since the advent of the "steering-from-a-distance" philosophy (see Maassen and van Vught, 1988), which recommended to enlarge institutional autonomy, to increase the flexibility of institutions, and to increase the diversity of the system. Based on analyses of the developmental plans of the universities (see also Potman et al., 1989), Maassen and Potman (1990b, p. 403) concluded that diversification had not been achieved: "The differences that already existed may continue, but innovations all seem to go into the direction of homogenization. As far as the development plans are concerned, the institutions have not succeeded in establishing meaningful and discriminating profiles. On the contrary, it seems likely that various homogenizing developments will emerge."

Maassen and Potman (1990a) believe that diversification was not enhanced because governmental policy is based on *competitive Sisomorphism*: institutions will, as a consequence of competition, search (and find) their own niche and develop their own mission and institutional profile. First, this scenario cannot work, according to the authors, because institutional actions are geared towards government and not towards competition with other institutions. Second, students' choices are based on factors other than the price and quality of the programs offered. Third, institutional policy has to take into account the norms and values of the academic profession. Fourth, the higher education field is highly structured, which has not changed substantially during the last decades.

The concept of *institutional isomorphism* (DiMaggio and Powell, 1983) helps explain the homogenization in Dutch higher education. Three mechanisms encourage institutional isomorphism: *coercive isomorphism*, *mimetic processes*, and *normative pressures*. Maassen and Potman (1990b) argue that these mechanisms are in force: e.g., in the new steering philosophy academic professionals play an important role in quality control (*normative pressures*), and government applies the same regulations to all universities (*coercive isomorphism*). To increase diversity, the government must formulate policies to take into account these mechanisms of isomorphism.

Conclusion

Looking at the conceptualizations (Table 2, page 80) used in the studies in this section, it is striking that the studies differ considerably in the use of the terms differentiation and diversification. Some studies on differentiation relate to specific entities emerging in the larger unit (e.g. Parsons and Platt, 1973), while other studies refer to an increase in the number of types of institutions within a system (e.g. Rhoades, 1983). There are also studies that conceptualize diversification as an increase of differences between (types of) institutions (e.g. Clark, 1983; Meek, 1991).

Notwithstanding the variety of conceptualizations, the research on differentiation of roles and functions shares certain themes. The developments of differentiation and dedifferentiation (or diversification and homogenization) can be understood best in the tension of power relationships between interest groups (Clark, 1978, 1983; Meek, 1991; Rhoades, 1983, 1990). In addition, the characteristics of the academic profession (Aldersley, 1995; Jencks and Riesman, 1968; Rhoades, 1983, 1990; Riesman, 1956), the policies and steering approach of the government (Maassen and Potman, 1990a,b; Meek, 1991; Rhoades, 1983, 1990) and the reaction of institutions, influence differentiation and dedifferentiation. The exception is the work of Parsons and Platt, who take an abstract and highly theoretical macro-level point of view, where individual actors and power groups have no place.

In general, the norms and values of the academic profession are considered as inhibiting factors for differentiation of roles and functions at the higher education system level.[3] Competition between institutions may stimulate differentiation by forcing institutions to look for their own niche in the market place (see also Birnbaum, 1983). Governmental policy may inhibit differentiation because it often restricts—by means of regulations—the scope of institutional actions; it forces institutions to resemble each other.

Some of the studies presented above challenge the rather simplistic conclu-

[3]Most studies are analytical, in the sense that the authors do not give empirical proof of this assertion. To give an example of empirical research supporting the assertion, Fairweather (1995) investigated faculty pay at different types of institutions, presuming that the different types show differences in reward structures (e.g. teaching versus research). His conclusion, however, is that faculty pay based on research and publications is predominant.

sions about the unidirectional influence of the factors like the academic profession, institutional competition, and governmental action on differentiation. Maassen and Potman (1990a), for instance, question the applicability of competition to Dutch higher education. Rhoades (1983, 1990), Skolnik (1986) and Meek (1991) maintain that the influence of the three factors are mediated by the context in which the power struggles between interest groups of higher education take place. Competition does not necessarily force institutions to develop distinctive profiles. When, for instance, research is highly valued by clients of higher education (industry, society at large), chances are that many institutions try to get, maintain, or enlarge their share in research activities instead of following the risky path of finding a specific (new) niche.

AN ASSESSMENT

In this section, I evaluate the achievements in the field of higher education with respect to describing, analyzing, and explaining differentiation processes and changes in the level of diversity. First, the differences between the reviewed studies will be discussed (including the strong and weak points). Second, the scientific yields of the studies presented above are set out. Therefore, the studies will be measured against the yardstick of the ideal research design, containing a theoretical framework, assumptions, hypotheses, operationalizations, testing, and feedback to the theoretical framework.

Three important differences between studies of differentiation of roles and functions and studies of classification (external diversity) exist: the attention paid to differentiation as a process versus diversity as a product, the object of study, and the theoretical considerations.

Most of the classification studies are static, relying on a snapshot in time. The studies picture the variety of institutions at a specific point in time (diversity as a product). Following the terminology of Table 2 (page 80), most classification studies relate to the variety of types within higher education systems, Birnbaum's (1983) is the only study that actually measures diversity in terms of the variety of types and the dispersion of institutions across these types. Typical of the studies is that institutions are clustered or classified using indicators such as enrollment characteristics, level of organizational efficiency, level of professional autonomy, and performance indicators. The variety of indicators used is striking. It highlights the fact that the classification results are largely determined by the choice of the indicators or variables (compare, e.g., the results of the Carnegie Foundation for the Advancement of Teaching studies with those of Birnbaum and Smart). In itself this is not problematic: as long as the researchers are clear on variables and methodology used, discussion is possible on the appropriateness of the chosen variables. Two of the studies show some dynamics: Birnbaum (1983) compares the level of diversity at two points in time, applying diversity indices such as the Gini-index and a Lorenz-curve analysis. Tight (1988) compares his results with previous classifications of British universities. The studies on differentiation of roles and functions pay much more attention to the process and the dynamic character of differentiation, where roles and functions change over time. In this respect, the conceptualizations come close to those of diversification (Table 2) and its opposite, homogenization.

A second difference concerns the object of study. The work of authors who focus on the roles and functions in higher education is relatively abstract. Some analyses refer to the level of the system (Rhoades, 1983, 1990; Meek, 1991), some focus on a specific type of institution, and a third set concentrates on roles and functions within systems of higher education as well as within institutions. Examples of the latter are Clark's (1983) and Parsons and Platt's (1973) work. Although Clark focuses on the disciplinary basic units of academic life, his argument of increasing differentiation (and fragmentation) is extended to other levels of the system: institutions, sectors, and tiers. The picture that emerges is evident: differentiation and diversity are ubiquitous characteristics of higher education (see also Clark, 1996; Trow, 1995). A striking feature of the studies on roles and functions is the lack of attention paid to defining and operationalizing the concepts diversity and differentiation (with few exceptions). In general it is clear that the authors refer to differences between sectors and institutions, but attention is hardly paid to important questions such as: in what respect and to what extent do the objects of study differ and, even more importantly, in what respect and to what extent do changes occur through time? On the contrary, many analyses draw on (perceptions of) developments and trends of homogenization and diversification. Classification studies, in contrast, define the objects unambiguously by focusing on institutions of—sectors of—higher education systems.

The third difference relates to the theoretical considerations. The authors in the section on roles and functions rely much more on theory than do authors of the classification studies. The theoretical considerations focus on power relationships, interest groups, and legitimacy of roles. Next to general consensus on the influence of the academic profession (limiting external diversity, promoting internal diversity, and differentiation), the studies disagree on the effects of governmental interference in higher education affairs. Some contend that (state) government inhibits diversity (by means of regulations, Birnbaum, 1983), others hypothesize that a certain amount of governmental regulation is necessary to maintain diversity (Rhoades, 1990). The studies also disagree with respect to the role of "market forces". Whereas some think that diversity will flourish when competition forces institutions to find their niche (Birnbaum, 1983), others think that competition strengthens imitative behavior and consequently leads to homogenization. These different perspectives are not necessarily at odds, for Birnbaum (1983) does not include market failures, whereas others (e.g. Maassen and Potman, 1990a) seriously consider these failures—in the Dutch context—such as students' choices not solely based on price and quality and the fact that competition is partly controlled by governmental regulation. The variety of theoretical assumptions and hypotheses must not be taken as a negative aspect of the studies. The existence of competing expectations allows for excluding explanations by empirical testing. However, some weaknesses appear especially concerning the explanatory power of the studies. Apart from the fact that operationalizations are not reported or are simply lacking, many studies get stuck in descriptive analyses of one case, which limits generalization. In addition, some generalizations are based on a limited number of cases: observing diversity in the United States higher education system in combination with weak (state) government intervention does not allow for a generalization such as "state interference limits diversity". Despite the potential usefulness of theory, the lack of operationalization of central notions and empirical testing seriously limit the results and conclusions of the studies on differentiation of roles and functions.

A general conclusion of the review is that the classification studies focusing on external diversity have their strong point regarding operationalization and methodology, but do not offer explanations for their findings. Contrarily, the studies on differentiation of roles and functions are relatively strong with respect to offering explanations. On other aspects of the ideal research design (operationalizations, definitions as well as the research methodology), these studies are less developed. Despite the criticism, some positive results must be stressed. First, the review brings to the fore the usefulness of distinguishing diversity (as a product) and differentiation (as a process). Second, many of the studies—although using different theoretical backgrounds—underscore the importance of legitimacy of roles, interest groups, dynamics, and power processes. The studies point at (possibly) relevant factors in processes of differentiation, such as governmental policies, the norms and values of the academic profession, and the forces of the "market". Third, combining the strong points of both classification studies (methodology, operationalizations) and differentiation of roles and functions studies (theoretical background) might offer opportunities to gain insight into factors explaining processes of differentiation and to increasing or decreasing diversity. This last point will be elaborated below. The main elements of a theoretical framework will be developed to investigate institutional diversity in higher education systems.

DEVELOPING A THEORETICAL FRAMEWORK

The issue of diversity (and/or homogeneity) of organizational forms has been studied by two different branches of organizational sociology, which propose competing explanations for the homogeneity or diversity of sets of interrelated organizations. The population ecology perspective (Hannan and Freeman, 1977) argues that initial variety within a population of organizations decreases through competition for scarce resources. This causes competitors to become more similar because the conditions of competition bring forth similar responses and, more important, because less optimal forms are selected out (competitive exclusion of forms using the same resources). Consequently, the population of organizations becomes more homogeneous by competitive isomorphism. Variety of organizational forms exists or emerges because different environments (with different resources, constraints, and opportunities) have different impacts on populations. Therefore, different environments lead to different populations of organizations. In this view, following a biological analogy, the homogenizing effects of competition are mediated by the "richness" of the environment. The diversity of organizational forms is considered proportional to the variety of resources that exist in the environment.

The new institutionalism in organizational analysis (DiMaggio and Powell, 1983; Meyer and Scott, 1983; Powell and DiMaggio, 1991) emphasizes the effect of institutional (instead of competitive) isomorphism on organizational forms. Hannan and Freeman's (1977: 936) question "Why are there so many kinds of organizations?" is countered with the question "[W]hy there is such startling homogeneity of organizational forms and practices...". (DiMaggio and Powell, 1983: 148). According to DiMaggio and Powell (1983), three (analytically distinguishable) mechanisms, namely *coercive*, *mimetic*, and *normative* isomorphism produce homogeneity within an organizational field, being those organizations that, in the aggregate, constitute a recognized area of institutional life (suppliers, consumers, regulating agencies, etc.). Contrary to population ecology, political power and legitimacy, rather than competition for resources, produces homogeneity of organizational structure. Coercive isomorphism results from pressure applied by other organizations on which the organization is dependent and by cultural expectations (e.g. governmental control, laws, technical requirements). Mimetic processes stem from uncertainty caused by poorly understood technologies, ambiguous goals, and the symbolic environment. The normative pressures stem primarily from professionalization. Two aspects of professionalization are relevant: the first is the homogenizing influence of formal university education, and the second is the growth and elaboration of professional networks.

Combining the Perspectives and Integrating Previous Research

Although both perspectives differ in their basic assumptions, recent publications suggest the existence of a good deal of common ground. Hannan and Freeman (1989) accept, for instance, that competition and constraints can have positive and negative effects on the diversity of forms, depending on specific circumstances. Additionally, empirical studies indicate that a combination of the two perspectives appear worthwhile. For instance, Barnett and Carroll (1993), Lomi (1995) and Singh, Tucker, and House (1986) point at institutional forces (legitimation) intermediating the effects of competition on population density. It must be stressed that the combination of the perspectives has mainly focused on growth and birth and death rates of populations of organizations. However, combining the two perspectives may also provide a viable basis for explaining homogenization or diversification in organizational fields.

The combination also provides the opportunity to bring together the somewhat disparate strands existing at present in the field of higher education studies that independently have touched upon some of the concepts and factors identified above. Examples in this respect are:

- The works of Aldersley (1995), Jencks and Riesman (1968), Maassen and Potman (1990b), Rhoades (1983, 1990) and Skolnik (1986) on the impact of professional values and professionalization in higher education, that to an extent appear to corroborate the basic argumentation underlying the concept of normative isomorphism.
- Birnbaum (1983), Neave (1996) and van Vught (1989) on the—presumed—effects of governmental regulation and the incorporation of more market-related elements (Birnbaum, 1983; Clark, 1983; van Vught, 1989) in higher education systems. In this, at present unsolved debate, empirical evidence is brought forward in support of (a) the notion that less governmental control over higher education will induce institutions to seek their own niche and thus increase diversity versus (b) that a certain degree of governmental control over institutions is necessary to safeguard differences between institutions, and thus maintain or even increase diversity. This debate is strongly related to the concepts of both competitive and coercive isomorphism.
- Several studies have also touched upon the concept of mimetic isomorphism. The works of Birnbaum (1983), Clark (1983), Jones (1996) and Meek (1991) indicate that copying behavior (whether in terms of emulating high rank institutions or in terms of adopting—seemingly—successful innovations at other institutions) might be an important factor decreasing diversity.

Hypotheses and Research Strategy

On the basis of the work undertaken in both population ecology and the new institutional perspective, four clusters of independent variables can be identified possibly explaining levels of diversity. From the population ecology perspective, it can be hypothesized that the more competition exists between institutions in the system, the more homogeneity is likely to occur, dependent on the degree of resource variety in the environment. In more operational terms, this implies that within this first cluster, attention will be paid to such variables as the degree of governmental regulation and planning, the degree of competition between institutions, and the number of resource bases for the institutions. The second cluster of variables relates to coercive isomorphism. Here, it can be hypothesized that the more higher education institutions are confronted by formal and informal pressures from other organizations on which they are dependent, the more homogeneous they will become. Again, in operational terms this implies that within this cluster attention will be paid to variables such as governmental policies—with particular emphasis on possible variety in impact at the institutional level, legal requirements, and the degree of institutional autonomy. The third cluster of variables deals with mimetic isomorphism. The more higher education institutions have to operate under conditions of increasing uncertainty, the more homogeneous they will become through modeling themselves on other institutions. Variables to be included are complexity of the environment and uncertainty regarding the continuous supply of critical resources. Finally, the fourth cluster will be derived from the concept of normative isomorphism. It can be hypothesized that the higher the level of professionalization within higher education institutions, the more homogeneous the system will become. Indicators to be used include ties with disciplinary organizations, the role of intermediate bodies, and the degree of stratification within the system. Note that some of the variables discussed fit different isomorphism mechanisms, this is due partly to DiMaggio and Powell's (1983) presentation of the mechanisms as being analytically—but not necessarily empirically—distinctive.

For testing the hypotheses, a combination of longitudinal and comparative research studies is proposed. The first approach allows for comparison within countries over time, which is necessary to address the issue of change in diversity. The second approach enables to assess the impact of (the combination of) different variables across countries. As will be clear from the macro-level approach, system-level characteristics are assumed to explain changes in the level of diversity. Consequently, variables will be measured by means of objective indicators at the level of the higher education systems. Note that the research agenda is rather ambitious. It does not imply that all variables or mechanisms should be involved in one study. Neither does the proposal disqualify micro-level research which highlights specific forms of isomorphism in specific settings (see Galaskiewicz and Wasserman, 1989; Mezias, 1990; Haveman, 1993 on isomorphism in other contexts than higher education). Below I give an example of the testing of a hypothesis concerning the influence of governmental regulation, based on an interpretation of my own research on programmatic diversity in the Dutch university sector (see Huisman, 1995, 1996, for an elaboration).

An Example: Programmatic Differences Among Dutch Universities

In the example the focus is on programmatic diversity in the Dutch university sector. It may seem that programmatic diversity relates to internal differences within universities. However, the example investigates differences and similarities between the program offerings of the universities and, consequently, relates to (changes in) external diversity.

The Dutch government has been an important actor in (decision making processes on) the provision of degree programs in the university sector. An analysis of the program offerings makes clear that through time (1975–1995) the programmatic diversity increased at the system level as well as between institutions. I maintain that (governmental) regulations and policies and the institutions' anticipation of the regulations were the most important factors increasing programmatic diversity. In terms of the theoretical framework, especially coercive pressure is considered as an important factor influencing the level of diversity.

Regulations

In the period 1960–1993, the Academic Statute laid down the general rules regarding the contents of program examinations and provides an overview of the officially recognized study programs. In general, universities could establish three kinds of programs. Academic programs—those already recognized—could be started after approval of the Minister of Education and Science. Experimental programs are programs not yet included in the Academic Statute. The Minister had to approve a proposal (after consulting with other concerned universities and advisory councils). Within ten years the program had to be included in the Statute. If the Minister agreed, the program achieved academic status. In other cases the experiment would end. Finally, the Statute allowed universities to establish free study programs, needing only the approval of the institution's decision making bodies.

Since 1993, institutions wanting to offer a new program must always register with the Minister. The registration must be accompanied by advice on whether to be included in the register or not by an advisory committee (its members are appointed by the Minister). An important criterion for inclusion is macro-level efficiency, i.e. the Minister and advisory committee take into account an efficient spread of (new) programs across the universities. Comparing the periods, the new regulations tighten the possibilities for institutions to start new programs by emphasizing macro efficiency and making it impossible to establish programs without external scrutiny by the Minister.

Policies

From the beginning of the 1980s, the Minister increasingly took into account the allocation of provisions in (dis)approving the establishment of new study programs. For instance, the Minister approved all proposals for a study program in informatics, under the condition that universities had to cooperate and had to stress different aspects of the field of informatics. From then on, a university was less sure of approval of new programs if such a program already existed at one of the other institutions, or if other universities were developing the same initiative. In the regulations this was made explicit: the revised 1982 Academic Statute obliged the institutions to justify the relationship of the new study program to other (more or less related) study programs.

Also in the 1980s, two governmental retrenchment operations led to the closing down and rearrangement of a fairly large number of study programs. The policies made clear that the government was able to intervene radically at the institutional level with respect to the provision of study programs.

The Impact of Governmental Policies and Regulation on Diversity

I hypothesize that the government had considerable influence with respect to the direction in which programmatic diversity has developed in the Netherlands. Institutions were, until the end of the 1970s, reasonably confident of approval of initiatives for new study programs. From the beginning of the 1980s this confidence disappeared. First, the regulations (Academic Statute, 1982) obliged institutions to justify the relationship of the new program to other—related—study programs, which forced the supporters to stress distinctive features of the program. Second, the retrenchment operations of the 1980s showed that the government was able to intervene at the institutional level with respect to the allocation of programs. Third, institutions noticed that in cases where different universities had developed similar new programs, the government stressed the need to cooperate and/or distribute tasks. Fourth, although the implementation of the new regulation in 1993 seemingly does not have consequences for the time period of the research, it well may be the case that institutions anticipated the implementation. The fact that new programs had to be approved by an advisory committee and the Minister under a stricter regime (macro efficiency), may have induced institutions to implement their new (free) programs just before the new regulations came into force.

I maintain therefore, that these developments "forced" the institutions to develop programs that differed from the already existing programs. From the early 1980s, government sent out signals that program duplication was to be contained (reallocation operation, 1993 regulations stressing macro efficiency). At the same time, the university sector was confronted by decreasing student numbers. Elsewhere (Huisman, 1995), I pointed out that starting new programs or specializations can be seen as a balancing operation to cope with the decreasing enrollments (and consequently decreasing governmental funding). These two developments, the governmental policies discouraging program duplication and the drive for institutions to establish new programs, are presumed to have led to an increase of institutional diversity: by implementing new programs not yet existing at other institutions, universities became less similar through time. Coercive pressure—applied by the government on which the universities are dependent, increased the level of diversity. This seems, at first sight, to contradict the assumption that coercive isomorphism leads to homogenization. The interference of the Dutch government in curriculum matters, however, should be interpreted as a necessary instrument preventing institutions from program duplication, which would be the consequence of normative (norms and values of the academic profession) and mimetic isomorphism (copying behavior under uncertain circumstances). To some extent governmental intervention in the Netherlands is comparable to intervention of state-wide governing boards on the provision of programs in United States public higher education systems (Morphew, 1996).

Measuring Institutional Diversification

As a measure of institutional similarity the percentages of study programs that each pair of universities have in common were calculated for the years 1975, 1980, 1985, 1990, and 1995. Similarity of programs is based on correspondence between names of programs. To dampen size effects the number of programs in common has been doubled and divided by the total number of programs of the two institutions. A hierarchical clustering technique was used, based on the minimum method (see Johnson, 1967; Meerling, 1981). The following figures show the results of this technique for the year 1975 and 1995 (figure 1 and 2).

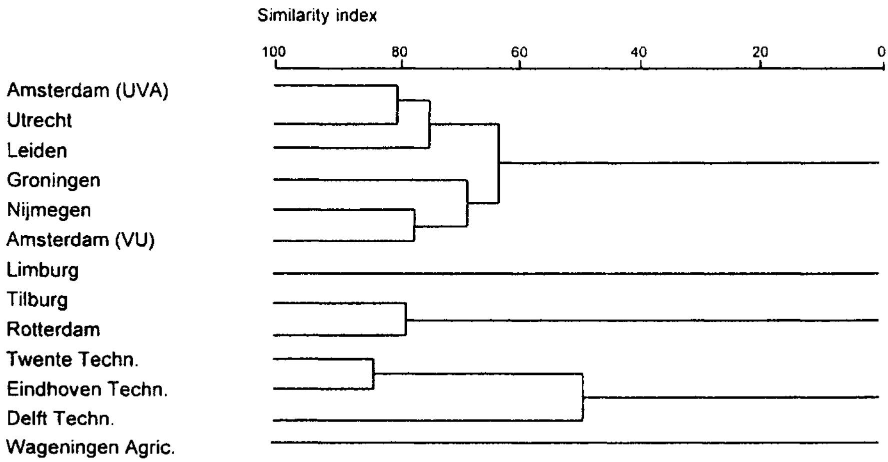


Figure 1: Cluster analysis of Dutch universities (program offerings 1975)

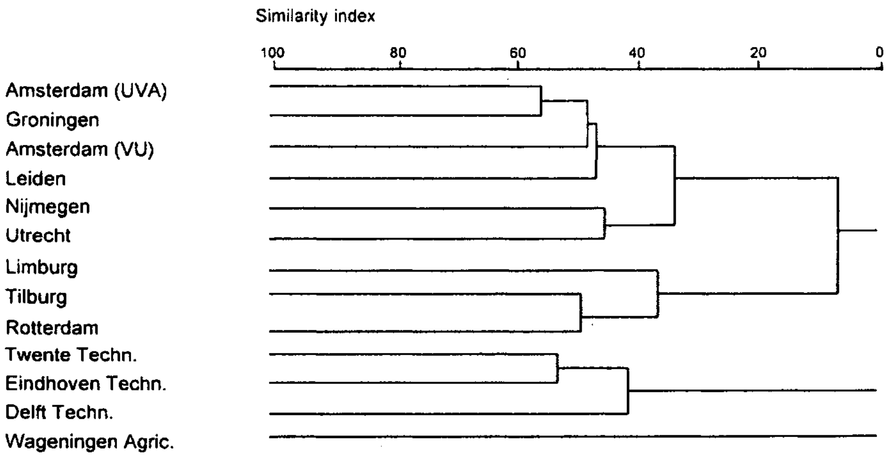


Figure 2: Cluster analysis of Dutch universities (program offerings 1995)

In general, the patterns are similar in both years with respect to the clusters emerging. The old and large universities form a rather homogeneous group. Tilburg and Rotterdam cluster together (two former economic hogescholen), as well as the three universities of technology. At the same time, it can be noticed that the clusters come into existence relatively earlier in 1975 (the similarity index ranges from 50 to 86 percent) than in 1995 (the index ranges from 8 to 57 percent). This means that the universities had—compared to 1995—on average more programs in common: through time the universities became less similar.[4]

At the same time it can be concluded that the clusters as a whole were more different in 1975 than in 1995 (five almost non-overlapping clusters versus three separate clusters). These results are partly inherent in the chosen method. Even when universities have only a small amount of programs in common, it is very likely that they will form a cluster, although the institutions may come together in a relatively late stage of the (cluster) analysis (i.e. the similarity index is low). The fact that only three clusters emerged in 1995 is due to the development of the relatively young University of Maastricht (established in 1974), which added a number of programs already existing at other universities. By 1995, the university is—figuratively—the bridge between the two small younger universities (Tilburg and Rotterdam) and the older/larger universities.

The hypothesis on the impact of governmental regulation and policies on the increase of institutional diversity (with respect to the program offerings) cannot be falsified. Governmental regulations and policies increased the differences between universities when looking at the study programs offered by the institutions through time. The effects of coercive mechanisms are demonstrated in the context of Dutch higher education.

CONCLUSION

Differentiation, diversity and related concepts like diversification, heterogenization, and homogenization are important issues in higher education practice and research. In this chapter an attempt was made—based on the use of the concepts in biology—to distinguish between differentiation, diversity, and diversification. A review of studies on external diversity on the one hand and differentiation of roles and functions on the other, revealed some noteworthy findings. The studies focusing on external diversity (i.e. differences between higher education institutions or types of institutions) show similar research designs. Using statistical techniques, typologies or classifications of higher education institutions are developed, which give an indication of the diversity of institutions at a specific point in time. Exceptionally, studies compare results over time. Contrary to the attention paid to methodological matters, the studies on external diversity hardly pay attention to explanations of the findings. The studies on differentiation of roles and functions are relatively strong with respect to offering explanations. In the theoretical framework power relationships, interest groups, and role legitimacy are key concepts. It must be stressed, however, that the theories differ with

[4]Data for the years 1980, 1985, and 1990—not presented here—confirm the gradual increase of institutional diversity with respect to the programs offered.

respect to the impact of specific factors (governmental regulation, market forces, academic norms and values) on the occurrence of processes of differentiation. Whereas the theoretical considerations must be mentioned as a strong point of the studies on differentiation of roles and functions, these studies often lack clear operationalizations of the key concepts and empirical testing of the hypotheses. Combining the fruitful aspects of the studies on external diversity (methodology, operationalizations) and the studies on differentiation of roles and functions (theoretical framework) might offer opportunities to gain (more) insight in factors promoting or inhibiting diversity and explanations of processes of differentiation.

An attempt has been made to develop a theoretical framework to investigate (and explain) institutional diversity and trends of homogenization or diversification, combining two branches of sociology: population ecology (Hannan and Freeman, 1977, 1989) and the new institutionalism in organizational analysis (DiMaggio and Powell, 1983; Meyer and Scott, 1983). Four interrelated mechanisms: competitive, coercive, normative, and mimetic isomorphism are assumed to affect the level of diversity of organization sets.

The Dutch example illustrated the effects of governmental regulation (coercive pressure) on institutional diversity (i.e. the differences between the program offerings of the institutions) in the Dutch university sector, and the appropriateness of the theoretical framework. The findings counter previous research and public opinion that maintain that governmental regulation and policies inhibit diversity (Birnbaum, 1983; Trow, 1995). Despite the positive results, some critical remarks are called for. The Dutch example has taken into account only one variable with possible influence on the level of diversity. Processes in present-day higher education systems are often too complex to be captured by such small sets of variables. Other variables should be included in the design to get a more complete picture of isomorphic change. Related to this, is the fact that only programmatic diversity is measured. It could well be the case that the Dutch institutions became more similar with respect to other characteristics (such as characteristics of the student body). Also, including other variables may uncover—unexpected—interacting effects and effects of intermediating variables. A stepwise elaboration of the model may be a fruitful research strategy. Notwithstanding the criticism, I expect the theoretical framework to be a sound and promising start to explain trends of homogenization and heterogenization. Currently, some of the hypotheses put forward are being tested, for instance, comparing the effects of governmental intervention on program duplication in a number of public higher education systems in the United States and the Dutch university sector (Huisman and Morphew, 1996). Also, using a comparative research design based on the theoretical notions outlined above, the effects of restructuring processes in Australian and Dutch higher education on systemic and institutional diversity are analyzed. Furthermore, a more practically oriented research project analyzes the effects of different steering strategies and policy instruments on programmatic diversity in both the Dutch university and the higher vocational education sector.

Results from these studies increase our knowledge on change processes in higher education as well as sustain governmental policies on diversity.

REFERENCES

Aldersley, S.F. (1995). 'Upward drift' is alive and well. Research/doctoral model still attractive to institutions. *Change* 27(5): 51-56.

Baldridge, J.V., Curtis, D.V., Ecker, G., and Riley, G.L. (1977). Diversity in higher education. Professional autonomy. *Journal of Higher Education* 48(4): 367-388.

Barnett, W.P., and Carroll, G.R. (1993), Organizational ecology approaches to institutions. In S.M. Lindenberg and H. Schreuder (eds.), *Interdisciplinary Perspectives on Organization Studies*. Oxford: Pergamon Press.

Biglan, A. (1973a). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology* 57(3): 195-203.

Biglan, A. (1973b). Relationships between subject matter characteristics and the structure and output of university departments. *Journal of Applied Psychology* 57(3): 204-213.

Birnbaum, R. (1983). *Maintaining Diversity in Higher Education*. San Francisco, CA: Jossey-Bass.

Blau, P.M. (1973). *The Organization of Academic Work*. New York, NY: Wiley.

Brint, S., and Karabel, J. (1989). *The Diverted Dream. Community Colleges and the Promise of Educational Opportunity in America, 1900-1985*. New York, NY: Oxford University Press.

Cameron, D.M. (1991). *More Than an Academic Question: Universities, Government and Public Policy in Canada*. Halifax, NS: Institute for Research on Public Policy.

Cameron, K.S. (1978). Measuring organizational effectiveness in institutions of higher education. *Administrative Science Quarterly* 23(4): 604-632.

Cameron, K.S. (1981). Domains of organizational effectiveness in colleges and universities. *Academy of Management Journal* 24(1): 25-47.

Carnegie Foundation for the Advancement of Teaching (1976 rev. edition). *A Classification of Institutions of Higher Education*. Princeton, NJ: CFAT.

Carnegie Foundation for the Advancement of Teaching (1987 rev. edition). *A Classification of Institutions of Higher Education*. Princeton, NJ: CFAT.

Carnegie Foundation for the Advancement of Teaching (1994 rev. edition). *A Classification of Institutions of Higher Education*. Princeton, NJ: CFAT.

Clark, B.R. (1978). Academic differentiation in national systems of higher education. *Comparative Education Review* 22(June): 242-258.

Clark, B.R. (1983). *The Higher Education System. Academic Organization in Cross-National Perspective*. Berkeley, CA: University of California Press.

Clark, B.R. (1996). Diversification of higher education: viability and change. In V.L. Meek, L.C.J. Goedegebuure, O. Kivinen and R. Rinne (eds.), *The Mockers and the Mocked; Comparative Perspectives on Differentiation, Convergence and Diversity in Higher Education*. Oxford: Pergamon.

Cowley, W.H., and Williams, D. (1991). *International and Historical Roots of American Higher Education*. New York, NY: Garland.

Department for Employment, Education and Training (1988). *Higher Education: A Policy Statement*. Canberra: Australian Government Publishing Service.

Department for Employment, Education and Training (1991). *Higher Education: Quality and Diversity in the 1990s*. Canberra: Australian Government Publishing Service.

DiMaggio, P.J., and Powell, W.W. (1983). The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. *American Sociological Review* 48(April): 147-160.

Dolton, P., and Makepeace, G. (1982). University typology: a contemporary analysis, *Higher Education Review* 14(3): 33-47.

Durkheim, E. (1893) 1964. *The Division of Labor in Society.* London: Free Press.

Eisenstadt, S.N. (1964). Social change, differentiation, and evolution. *American Sociological Review* 29: 375-386.

Fairweather, J.S. (1995). Myths and realities of academic labor markets, *Economics of Education Review* 14(2): 179-192.

Fox, W. (1994). Higher education policy in California. In L.C.J. Goedegebuure, F. Kaiser, P.A.M. Maassen, V.L. Meek, F.A. van Vught and E. de Weert (eds.), *Higher Education Policy. An International Comparative Perspective.* Oxford: Pergamon Press.

Galaskiewicz, J., and Wasserman, S. (1989). Mimetic processes within an interorganizational field: an empirical test. *Administrative Science Quarterly* 34(3): 454-479.

Geiger, R.L. (1986). *To Advance Knowledge: The Growth of American Research Universities, 1900-1940.* New York, NY: Oxford University Press.

Goedegebuure, L.C.J., Kaiser, F., Maassen, P.A.M., Meek, V.L., van Vught, F.A., and de Weert, E. (1994). International perspectives on trends and issues in higher education policy. In L.C.J. Goedegebuure, F. Kaiser, P.A.M. Maassen, V.L. Meek, F.A. van Vught and E. de Weert (eds.). *Higher Education Policy. An International Comparative Perspective.* Oxford: Pergamon Press.

Goedegebuure, L.C.J., Lysons, A., and Meek, V.L. (1993). Diversity in Australian higher education? *Higher Education* 25(4): 395-410.

Goedegebuure, L.C.J., and Westerheijden, D.F. (1991). Changing balances in Dutch higher education. *Higher Education* 21(4): 495-520.

Gruber, K.H. (1993). *Breaking the (University) Mould? The Beginning of Higher Education Diversification in Austria.* Paper prepared for the conference on Diversification in Higher Education, University of Turku, Finland, 13-15 August 1993.

Hannan, M.T., and Freeman, J. (1977). The population ecology of organizations, *American Journal of Sociology* 82(5): 929-964.

Hannan, M.T., and Freeman, J. (1989). *Organizational Ecology.* Cambridge: Harvard University Press.

Haveman, H.A. (1993). Follow the leader: Mimetic isomorphism and entry into new markets. *Administrative Science Quarterly* 38(4): 593-627.

Huisman, J. (1995). *Differentiation, Diversity and Dependency in Higher Education. A Theoretical and Empirical Analysis.* Utrecht: Lemma.

Huisman, J. (1996). Diversity in the Netherlands. In V.L. Meek, L.C.J. Goedegebuure, O. Kivinen and R. Rinne (eds.), *The Mockers and the Mocked; Comparative Perspectives on Differentiation, Convergence and Diversity in Higher Education.* Oxford: Pergamon.

Huisman, J., and Morphew, C.C. (1996). *Accounting for Program Diversity: Perspectives on American and Dutch Higher Education.* Paper presented at the 21st ASHE Conference, Memphis, USA, 30 October 1996.

Huston, M.A. (1994). *Biological Diversity. The Coexistence of Species on Chancing Landscapes.* Cambridge: Cambridge University Press.

Jencks, C., and Riesman, D. (1968). *The Academic Revolution.* Garden City, NY: Doubleday.

Johnson, S.C. (1967). Hierarchical clustering schemes. *Psychometrika* 32(3): 241-154.

Jones, G.A. (1996). Diversity within a decentralized higher education system: the case of Canada. In V.L. Meek, L.C.J. Goedegebuure, O. Kivinen and R. Rinne (eds.), *The Mockers and the Mocked; Comparative Perspectives on Differentiation, Convergence and Diversity in Higher Education.* Oxford: Pergamon.

King, J. (1970). The typology of universities, *Higher Education Review* 2(3): 52-61.

Lomi, A. (1995). The population ecology of organizational founding: location dependence and unobserved heterogeneity. *Administrative Science Quarterly* 40(1): 111-144.

Ludwig, J.A., and Reynolds, J.F. (1988). *Statistical Ecology. A Primer on Methods and Computing*. New York, NY: Wiley & Sons.

Lysons, A. (1990a). Taxonomies of higher educational institutions predicted from organizational climate, *Research in Higher Education* 31(2): 115-128.

Lysons, A. (1990b). Dimensions and domains of organizational effectiveness in Australian higher education. *Higher Education* 20(3): 287-300.

Maassen, P.A.M., and Potman, H.P. (1990a). Strategievorming in het hoger onderwijs: problemen bij het streven naar meer differentiatie van universiteiten. *Beleidswetenschap* 4(2): 112-130.

Maassen, P.A.M., and Potman, H.P. (1990b). Strategic decision making in higher education. An analysis of the new planning system in Dutch higher education. *Higher Education* 20(4): 393-410.

Maassen, P.A.M., and van Vught, F.A. (1988). An intriguing Janus-head: the two faces of the new governmental strategy for higher education in the Netherlands. *European Journal of Education* 23(1/2): 65-76.

Marcus, L.R. (1994). Diversity and its discontents. *Review of Higher Education* 17(3): 225-240.

Meek, V.L. (1991). The transformation of Australian higher education from binary to unitary system. *Higher Education* 21(4): 461-494.

Meerling (1981). *Methoden en technieken van psychologisch onderzoek. Deel 2: data-analyse en psychometrie*. Meppel: Boom.

Meyer, J.W., and Scott, W.R. (eds. 1983). *Organizational Environments. Ritual and Rationality.* London: Sage.

Mezias, S.J. (1990). An institutional model of organizational practice: Financial reporting at the Fortune 200. *Administrative Science Quarterly* 35(3): 431-457.

Ministry of Education (1992). *Developments in Education 1990-1992. Finland.* Helsinki: Ministry of Education.

Morphew, C.C. (1996). *Statewide Governing Boards and Program Duplication: A Longitudinal Study of Seven Public Systems of Higher Education.* Stanford University: unpublished dissertation.

Neave, G. (1979). Academic drift: some views from Europe. *Studies in Higher Education* 4(2): 143-159.

Neave, G. (1983). The dynamics of integration in non-integrated systems of higher education in Western Europe. In H. Hermanns, U. Teichler and H. Wasser (eds.), *The Compleat University.* Cambridge, MA: Schenkman.

Neave, G. (1996). Homogenization, integration and convergence: The Cheshire cat of higher education analysis. In V.L. Meek, L.C.J. Goedegebuure, O. Kivinen and R. Rinne (eds.), *The Mockers and the Mocked; Comparative Perspectives on Differentiation, Convergence and Diversity in Higher Education.* Oxford: Pergamon.

Organization for Economic Co-operation and Development (OECD)(1990). *Higher Education in California. Reviews of National Policies for Education.* Paris: OECD.

Parsons, T. (1966). *Societies: Evolutionary and Comparative Perspectives.* Englewood Cliffs, NJ: Prentice-Hall.

Parsons, T. (1978). *Action Theory and the Human Condition.* New York, NY: Free Press.

Parsons, T. and Platt, G.M. (1973). *The American University.* Cambridge, MA: Harvard University Press.

Patil, C.P. and Taillie, C. (1982). Diversity as a concept and its measurement. *Journal of the American Statistical Association* 77(September): 548-561.

Pielou, E.C. (1977). *Mathematical Ecology.* New York, NY: Wiley & Sons.

Potman, H., Binsbergen, P.A., and van Vught, F.A. (1989). *Geplande plannen? Onderzoek naar de doorwerking van het HOOP binnen de ontwikkelingsplannen van univer-*

siteiten en hogescholen in Nederland. Enschede: Centrum voor Studies van het Hoger Onderwijsbeleid.

Powell. W.W. and DiMaggio, P.J. (eds. 1991). *The New Institutionalism in Organizational Analysis.* Chicago, IL: University of Chicago Press

Rhoades, G. (1983). Conflicting interests in higher education. *American Journal of Education* 91(3): 283-327.

Rhoades, G. (1990). Political competition and differentiation in higher education. In J.C. Alexander and P. Colomy (eds.), *Differentiation Theory and Social Change. Comparative and Historical Perspectives.* New York: Columbia University Press.

Riesman, D. (1956). *Constraint and Variety in American Education.* Lincoln, NE: University of Nebraska Press.

Rueschemeyer, D. (1977). Structural differentiation, efficiency, and power. *American Journal of Sociology* 83(1): 1-25.

Singh, J.V., Tucker, D.J., and House, R.J. (1986). Organizational legitimacy and the liability of newness. *Administrative Science Quarterly* 31(2): 171-193.

Skolnik, M.L. (1986). Diversity in higher education: The Canadian case. *Higher Education in Europe* 11(2): 19-32.

Smart, J.C. (1978). Diversity of academic organizations. Faculty incentives. *Journal of Higher Education* 49(5): 403-419.

Smart, J.C., and Elton, C. (1982). Validation of the Biglan model. *Research in Higher Education* 17(3): 213-229.

Smelser, N.J. (1959). *Social Change in the Industrial Revolution.* Chicago, IL: University of Chicago Press.

Smith, D.G., Morrison, D.E, and Wolf, L.E. (1994a). College as a gendered experience. An empirical analysis using multiple lenses. *Journal of Higher Education* 65(6): 696-725.

Smith, D.G., Wolf. L.E., and Levitan, T. (eds. 1994b). Studying diversity in higher education. *New Directions for Institutional Research* 81.

Spencer, H. (1898). *Principles of Sociology.* New York, NY: Appleton.

Stadtman, V.A. (1980). *Academic Adaptations. Higher Education Prepares for the 1980s and 1990s.* San Francisco, CA: Jossey-Bass.

Stanley, G., and Reynolds, P. (1994). Similarity grouping of Australian universities. *Higher Education* 27(3): 359-366.

Stinchcombe, A.L. (1965). Social structure and organizations. In J.G. March (ed.), *Handbook of Organizations.* Chicago, IL: Rand McNally.

Stoecker, J.L. (1993). The Biglan classification revisited. *Research in Higher Education* 34(4): 451-464.

Tight, M. (1988). Institutional typologies. *Higher Education Review* 20(3): 27-51.

Trow, M. (1995). *Reflections on Diversity in Higher Education.* Paper presented at a colloquium at the Technische Hochschule Zürich, Switzerland, 31 August 1995.

van den Bijtel, J.A.L. (1988). *Analyse van het onderwijsaanbod van eerste-fase-opleidingen.* Evaluatie-onderzoek WTFS deelrapport 4. Amsterdam: SCO.

van Vught, F. A. (ed.) (1989) *Governmental Strategies and Innovation in Higher Education.* London: Jessica Kingsley.

Whitley, R. (1984). *The Intellectual and Social Organization of the Sciences.* Oxford: Clarendon Press.

Zammuto, R.F. (1984). Are the liberal colleges an endangered species? *Journal of Higher Education* 55(2): 184-211.

Reviewing and Rethinking Administrative Costs

Gary Rhoades
Center for the Study of Higher Education, University of Arizona

If much has been written and said about administrative costs, little is empirically known. State agencies and many faculty have criticized increased administrative costs. Many scholars have pointed to various factors such as increased governmental regulations and decreased faculty involvement with students as necessitating increased administrative costs. However, the terms of the debate are inadequate to a useful analysis of the phenomenon in ways that might inform practice and enhance our scholarly understanding of the ongoing restructuring of higher education's work force.

In reviewing the literature on administrative costs, I argue that we need to refine and rethink the expenditure and personnel categories that we currently use. We need to refine categories in the sense of disaggregating them. In order to fully understand and explain administrative costs, we must be able to target precisely where those costs are being incurred. We need to rethink categories in the sense of overcoming the prevailing dichotomy of administration versus faculty. Such a conceptualization blinds us to the fundamental restructuring and restratification of professional and non-professional work forces in colleges and universities. It also blinds us to the changing nature/functions of non-faculty personnel, which help explain expenditure and personnel shifts in "administrative costs." Finally, conceptualizing "administration" in terms of "costs" inhibits our understanding of non-faculty personnel's productivity, their contribution to the production work of the academy.

In this chapter, I first develop the analytical points identified above. I then review the administrative costs literature in light of these points. Finally, I provide two sets of empirical data to clarify and highlight the significance of the two analytical points I am making—the need to disaggregate and to reconceptualize non-faculty personnel.

Refining Expenditure and Personnel Categories of administrative costs

Too often, commentary, scholarship, and policy deliberations fail to disaggregate the "administration" in administrative costs. Yet, patterns of costs vary consider-

J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*

ably within the broad, non-faculty, residual category of "administration." I suggest three dimensions of disaggregation.

Expenditures on administration can be disaggregated to the general type of administration. National data sources report expenditures in the general categories of student services, institutional support, academic support, public relations. Of course, as with any general category, the specific units that fall into one functional division or another vary by place and over time. From one institution to another, the range of operational units that fall under student services (or other divisions) varies. So, too, over time, often within the same institution, the range of operational units that fall under academic support (or other divisions) also varies. Nevertheless, the general functional divisions are critical to understanding administrative costs.

Patterns of administrative costs vary not only among these functional divisions, but within them. In analyzing costs we need to disaggregate to the level of operating units and divisions within each of the broader categories. The greatest resource shifts may be taking place not between but within these functional divisions. For example, student services may not be losing much of its expenditure share (of institutional expenditures), but within student services, units devoted to counseling services may have reduced expenditures relative to units devoted to placement (Chavez, 1996). Similarly, academic support, or institutional support, may be increasing expenditure shares (of institutional expenditures), but within these divisions, units such as affirmative action (which is generally in one or the other division) may have reduced expenditures. The story of administrative expenditures is in part one of internal reprioritizing and restratification of units' resource shares.

A second dimension of disaggregation is personnel categories, which can be separated into broad divisions beyond the simple classification of "administration." We need to move beyond grouping all non-faculty personnel into the residual category of "administrators," which is too often and too quickly translated into "bureaucrats." (As the old joke goes, remembered to me by Harold Weschler, there's the faculty and the "t'aint (it ain't) faculty.") In the National Center for Education Statistics (NCES) data, which utilize the Equal Employment Opportunity Commission's (EEOC) seven job categories, there are two classes of professional workers that are not faculty: administrative, executive, and managerial; and support professional. The former refers to personnel with primary and major managerial responsibilities. The latter refers to personnel with college degrees who serve more in a support capacity. The distinction is not unlike the classic organizational distinction between line and staff positions. However, it is also simply a matter of positional hierarchy. For example, Directors (e.g., of Personnel, Physical Plant, Alumni Affairs, or of Admissions) are classified as "support professionals" although personnel in these positions engage in managerial activities as much as do the Deans, Associate Vice-Presidents, and Senior Vice-Presidents above them. In addition to these classes of professional personnel there are four classes of "non-professional" personnel: technical and paraprofessional; clerical and secretarial; skilled crafts; and service/maintenance. Again, these are not just functional or education based divisions, they are a matter of positional hierarchy.

Just as one can disaggregate within expenditure categories, so one can disaggregate within personnel categories. For example, there are executive positions in student services, in academic support, in institutional support, and in public service. More than that, however, the category of support professional ranges across various professions and very different types of personnel. Of course, many of these are related to the functional divisions within student services, academic support, institutional support, and public relations. If we are to understand the cause and consequences of increased "administrative" costs, we must be able to specify those costs to the disaggregated level of accountants, counselors, technology transfer and licensing specialists, computer and systems analysts, and a range of other professionals, within particular functional divisions of administration. In addition, we need to specify those costs to the disaggregated level of such professional personnel, or to non-faculty personnel.

A third dimension of disaggregation is to go beyond the focus on central administration (and support staff). There are large numbers of executive and support and non-professional personnel below and outside of central administration. There are large numbers of such personnel in academic units. For example, academic colleges, departments, centers/interdisciplinary units, and research units have executive, support, and non-professional personnel. Many academic deans, particularly in professional schools, not only have associate deans for academic affairs and/or for student services, they also have support professionals in development, business (and sometimes research), and computer support. Most academic units in science, math, engineering, and health sciences have various support professional and non-professional personnel such as lab technicians and grants writers. Administrative costs extend below and outside of the central offices of campuses (indeed, they may even be counted as "instructional" costs).

Going beyond the central administration of a campus need not only mean looking more carefully within the campus. It may also mean looking above the central administration. In the public sector, most campuses are part of larger organizational systems, whether they be state systems, metropolitan districts, or multiple campuses of a single institution. Administrative costs can be incurred at these supra-campus levels. For example, in California, the University of California, the California State Universities, and the Community Colleges each have systemwide administrations that are massive in size. In the community college sector, there are also many large urban, metropolitan districts that have large central administrative units that sit above the individual colleges. In both two- and four-year sectors there are multicampus institutions that have administrative components that not only oversee each campus, but that cross-cut the campuses. Finally, many states have governing boards of trustees that themselves have increasingly large staffs. In most of the higher education literature, these boards are conceptualized as being separate from the institutions. Yet they are formally at the top of the governance hierarchy and can also be conceptualized not simply as a cause of increased administrative costs (through increased regulations and requests for information), but as an entities that themselves directly incur administrative costs. Indeed, in some states, such costs have been a significant political issue, a source of criticism by legislatures and/or governors, and the basis of efforts to disband such superboards.

Rethinking the Dichotomy of Administration Versus Faculty

More than this refinement and disaggregation, though, we need to rethink the terms "administration" and "costs." The discussion/debate/dispute and even the scholarship about administrative costs generally suggest a dichotomy of faculty versus administration. This simple separation is built into our everyday language, our work, our professional affiliations, our conceptualizations, and into our theories about colleges and universities.

No doubt, many of us have heard faculty colleagues equate administrators with "bureaucrats," contrasting them with the professionals (faculty) who do the central and "real" work of the organization. Likewise, many of us have heard administrative colleagues discuss the time demands and full workdays of their lives, contrasting them with the free Fridays, summers, and early exits from campus of faculty. Perhaps many of us have heard the joke told by senior faculty that they can remember a time when there were more "janitors" than vice-presidents. College and university hallways and offices echo with faculty complaints about administrators' salaries, and with administrators' complaints about many unproductive faculty who draw salaries but hardly ever come "to work" (that is to their designated workplaces, their offices). The lines are drawn between faculty and administration in our discourse.

In Higher Education programs, in talking about our students and their career destinations, we often distinguish between the majority who are going into administration and those relatively few who are pursuing faculty positions (also, in our research, see Townsend and Mason, 1990). In describing our colleagues we often refer to the mix of those who have administrative experience and those who have "only" been faculty, never having participated directly in the "real world" of practice. And we talk about peoples' transition from administration to faculty (and vice versa) as if it involves resocialization into a fundamentally different world. In teaching our classes, we seek to connect the academic world of scholarship to the world of administrative practice. In our work lives we invoke these dichotomous categories of faculty and administration.

The birth of our professional association, the Association for the Study of Higher Education, was grounded in the separation of professors who study higher education from top administrators who manage higher education. Within that association we now hear administrators expressing a sense that they do not feel entirely welcome and faculty expressing a sense that their work is not sufficiently relevant to or utilized in the world of administrative practice. We now work to bring together administrators and faculty in a community of mutual respect. Yet, if we share our ASHE affiliation, there are often vast differences in our other professional affiliations. Those of us who are lifetime faculty are more likely to belong to disciplinary and scholarly associations in which attendees dress more casually, "like faculty" (more sports coats and slacks, blouses and pants, jeans, sweaters). Those of us who are currently administrators (or "administrator identified") are more likely to belong to institutional associations in which attendees dress more formally, "like administrators" (more suits, for both men and women). In our travels (travel budgets) and professional affiliations, one can distinguish administrative and faculty circles, which may at times overlap but are distinct, as in a Venn diagram.

Finally, concepts and theories generated and utilized in our scholarship (and in our professional practice) are grounded in the separation between administrators and faculty. For example, in characterizing our organizations' structure and governance (and culture), traditionally we have counterposed the bureaucratic and professional against each other. Thus, we have the classic analyses of professionals working within the structure of bureaucratic organizations, even to the point of analyzing how professionals and administrators occupy different buildings on campus (Lunsford, 1970). Our models of governance contrast the power of position (managerial) versus the power of expertise (professional) (Blau and Scott, 1962; Goode, 1957; Parsons, 1954; Scott, 1966). We have the classic separation of domains of responsibility in governance—faculty (professionals) control the curriculum), and administrators (bureaucrats) control the finances (Mortimer and McConnell, 1978). We speak of a bureaucratic mentality (and culture) that seeks to control professionals and minimize discretion and professional judgment with formal policies and rules. More recently, institutional theory characterizes the relationship differently (Meyer and Rowan, 1977; Powell and DiMaggio, 1991). A distinction is still drawn between formal bureaucratic structure and actual professional activity. However, in this model the bureaucratic is seen as buffering the professional from the outside world. In some of our theories of resource allocation, we distinguish between academic and non-academic units, as if this is the major fault line between what is "central" and what is "peripheral" or "ancillary" (Hackman 1985). Indeed, in some cases, we speak of institutions having held the academic side of the institution relatively harmless in difficult fiscal times by cutting support and administrative units and by distinguishing between "infrastructure" and academic programs. In comparing salaries, it is more common (not just in our discourse, but in our scholarship as well) to focus more on differences between academics and administrators than on internal differentiation within those categories (Hansen and Guidugli, 1990; c.f. Slaughter and Rhoades, 1996).

I am not arguing that there are no differences between administration and faculty. I am arguing that the differences are not always so clear (in our everyday work lives) as we make out. For example, some faculty do a considerable amount of administrative work. Among other tasks, they sit on committees that draft various sorts of policies (e.g., strategic plans) and that make various sorts of decisions (e.g., personnel, admissions). Also, many researchers, particularly in the sciences, administer grants and hire and supervise personnel (e.g., lab technicians, postdoctoral scholars). Some teaching faculty oversee teaching assistants and coordinate classes. So, too, some administrators teach classes, some do research, and most perform some sort of community and professional service. Moreover, students in Higher Education do not simply go into either teaching or administration. Many seek to do both, sometimes at different points in their careers. In some cases, that means first going into administration and then teaching and/or becoming full-time faculty at a later point in their careers. In some other cases, that means first going into a faculty position on the way to becoming an academic administrator. At any rate, many faculty in Higher Education programs seek to train "scholar administrators" or "reflexive practitioners," or scholars who address issues relevant to practitioners and policymakers in the field. In our research, too, the lines are not as distinct as we sometimes make out. For example, how do we classify certain positions? In various studies, department heads have been classified as administrators or as faculty (Fairweather, 1993; but most studies of administrative costs classify department heads as faculty). Indeed, the nature of the position may vary by the type of department and institution. My point is that the differences are not so clear as we oftentimes assume.

I am also arguing that in characterizing administration and faculty as dichotomous categories we are blinded to substantial changes in the higher education work force. That is particularly true in the case of administrators, who get studied far less than faculty. It is time we recognized that non-faculty personnel, particularly in recent years, have become more directly involved in the "production work" of the academy—in producing students and research. Such changes are in one sense reducing the differences between administration and faculty. In another sense, such changes are contributing to the development of new fault lines and of increasing differences among higher education personnel as we restructure our colleges and universities. Dichotomizing administrators and faculty blinds us to the changing nature and functions of non-faculty personnel, to the restructuring of the work force in higher education. The classic view, and the faculty view, is that faculty are THE production workers on campus. Recall the often-told, and probably apocryphal, story of the Nobel Prize-winning faculty member at Columbia University rising and saying to the new President, who had been referring to faculty as "employees," "We are not the employees of Columbia, we are Columbia University." Faculty believe that they are the ones who create value, whether in research or in education. It is time to rethink that view.

Faculty are no longer, if they ever were, the only "production workers" on campus. Non-faculty personnel have become increasingly involved in the production of students and research. For example, there are "administrative" personnel and functions directly involved in the education and graduation of students, from the area of computer support to the co-curricular program to honors centers to tutoring, skills development, and advising. Indeed, it may be that an institution can contribute more to enhancing students' overall educational experience, and to improved persistence and graduation rates, not simply by hiring more faculty, but by hiring more non-faculty personnel. That is an empirical question, and one that has yet to be explored. Similarly, there are "administrative" personnel and functions directly involved in the production of research, from the offices of vice-presidents for research to sponsored projects personnel to grants writers and administrators. Again, an institution may enhance its research and grants production not simply by hiring more faculty to do research, but by hiring more non-faculty personnel.

Moreover, as institutions have increasingly emphasized the generation of revenues, non-faculty personnel have become increasingly central in that production function. There are "administrative" personnel and functions directly involved in generating revenues, from fund raisers to managers of university/business partnerships (e.g., research parks) to technology transfer professionals. An institution may enhance its fiscal self-sufficiency by reducing the percentage of faculty on campus, by hiring increasingly more non-faculty personnel. That is another empirical issue that has yet to be explored in the literature.

Dichotomizing faculty and administrators blinds us to the production role of non-faculty personnel. As a result, scholars are less likely to examine the productivity of these new "production workers." Non-faculty personnel are not just "costs," expenditures that have no yield. They, too, generate outputs and create value. As non-faculty personnel play increasingly important roles in producing students, research, and revenues (not to mention outreach/service), it is increasingly important to examine how effective and efficient they are in these roles, just as increased attention is being directed to faculty's effectiveness and productivity.

Reconceptualizing the role of non-faculty personnel leads us to address another issue about non-academic units, analogous to one that gets addressed in examining academic programs. As colleges and universities restructure, how is the hierarchy among units restratified? On the academic side, it is clear that some programs are more likely to get downsized and/or cut than others, and that the criteria for allocating resources (including new faculty lines) is related to programs' ability to generate external revenues and to their perceived "closeness to the market" (Hackman, 1985; Slaughter, 1993; Slaughter and Rhoades, 1996). So, too, non-academic units are not a monolith: they are internally stratified, and are increasingly so. Too often, practitioners speak of sacrificing the support side of the institution to protect the academic, blinding us to patterns of restratification among support units.

Reviewing the Literature: Administrative Costs

Given the careers of ASHE members, and the occupational destinations of their students, we know surprisingly little about non-faculty professionals on campus. There is little scholarship on such personnel. There is even less on the costs and productivity of employees other than faculty. As one study of "cost containment in higher education" put it:

> Emphasis in the literature on university administration is primarily placed on how the services offered may be improved, to a lesser extent on how greater efficiency may be obtained, and to an even lesser extent on how advances may be made in cost-effectiveness. (Simpson, 1991, p.207)

In canvassing the top journals in the field of Higher Education, I found few articles addressing the subject of administrative costs. From 1990 to the present, I found two articles in the *Journal of Higher Education*, none in the *Review of Higher Education*, none in *Higher Education* (there is one special issue, v. 25, n.3, in 1993, on Total Quality Management, but the focus is on process efficiency and morale, not on administrative costs), one in *Research in Higher Education*, and one in *Change*. Even in the *CUPA Journal* (of the College and University Personnel Association), which focuses on administrative and support staff, articles on this topic are notably absent (there is one such article). By contrast, there are several articles dealing with issues such as "Is your human resources department staffed adequately?" (Bouchard, Davidson, and Fortunato, 1990) and "Toward establishing salary benchmarks for college and university administrators" (Summerville and Ridley, 1992).

Although overlooked in higher education journals, administrative costs have become a significant political issue in the 1990s. Such costs have been raised as an issue by major faculty unions. For example, an entire issue of *Academe* (November/December, 1991) the journal of the American Association of University Professors, was devoted to "Administrative Bloat" (see articles by Andersen, Bergmann, and Halfond). Administrative costs have also been raised as an issue by task forces and boards in several states, including Arizona, Iowa, and Massachusetts, among others. All the more surprising, then, that higher education scholars have largely overlooked the issue.

Outside of the higher education journals, there is also very little work on the subject. In the late 1960s and early 1970s a few studies emerged in sociology and organization theory journals as scholars began to deal with the correlates and effects of organizational size and complexity. Most of these studies address Parkinson's (1957) observation that with increasing size, the proportion of administrators to production workers increases (also see Caplow, 1957). Several studies found that administrative staff are a decreasing function of an organization's size and complexity (Hendershot and James, 1972; Indik, 1964). Some of that work applied to higher education (Hawley, Boland, and Boland, 1965), although the findings in that case were mixed: complexity contributed to an increased proportion of administrators if size (measured as number of full-time faculty) was held constant (in itself, size contributed to a reduced proportion of administrators citations). Indeed, subsequent studies of these issues generally began to focus on interactive effects (Klatzky, 1970).

Interestingly, in subsequent years, administrative costs dropped out of the literature as a significant problem and focus. Indeed, from 1990 to the present, the leading journal in organizational theory, *Administrative Science Quarterly*, had no articles on administrative costs.

Why do we pay so little attention, in relative terms, to administrators and their costs? It might have something to do with the nature and function of our field and units. For all the concern about gaps between theory and practice, Higher Education as a field of study is oriented to issues that are relevant to managerial practice—more than, for instance, to a study of grassroots social action and change. That pattern is not unlike what holds for organization theory, which Higher Education scholars draw on considerably. In the 1960s, at around the same time Higher Education programs began to grow, organization theory separated itself out from industrial sociology, and began to focus not on the study of work, but on "seeking general principles of organizing" (Barley, 1996, p.405). The orientation of the work became much more managerial, concentrating on questions surrounding management rather than on issues related to the work and position of employees. Similarly, Higher Education units are advisory more to management, to central administration, than to faculty unions, associations, or organizations on campus such as senates. Likewise, such programs are more advisory to student affairs professionals than to students themselves, including various subdivisions and organizations of students. We are hardly likely to "bite the hand" we advise and that hires our students.

Yet, that does not explain why there is relatively little research on other aspects of administration. In my view, a large part of the explanation for the field's failure to adequately address the subject of administrators generally, and of administrative costs in particular, lies in our (mis)conceptions about the higher education work force, in the categories that frame our thinking. The terminology we use of "administration" and "support" suggests employees who are ancillary to the primary professionals and functions of the academy. (Indeed, that view is embedded in some of our theories—e.g., see Hackman, 1985, on "centrality.") The terminology also suggests not just a minimal (or subordinate) role, but limited size.

However, as several observers have noted and documented, professionals other than faculty have accounted for most of the growth in professional positions on campuses. From 1975 to 1985, two categories of professional employees—administrative/managerial/executive and other professional—experienced 18 and 61 percent growth respectively. By contrast, faculty experienced only 6 percent growth. From 1985 to 1990, increases for the three categories of executive, support professional, and faculty were 14, 28, and 9 percent respectively (Grassmuck, 1990, 1991). By 1991, faculty accounted for only 51 percent of the professional work force, and less than one-third (29 percent) of the total work force in higher education. In fact, in the period from 1987 to 1991, faculty experienced a 4.9 percent *decline* in numbers, the only job category with a decline except for service workers, which experienced a 1.3 percent decline (Montgomery and Lewis, 1995).

As of this writing, we lack national data to track the most recent five-year period, between 1991 and 1996. This period is particularly important in that it includes a period during which for the first time since the Great Depression, state appropriations in many states declined in absolute terms. For public institutions in particular, the early 1990s were a period of cost containment and budget cuts. Has the growth of administrative and support positions continued to increase relative to the academic side?

Many college and university central administrators claim to have protected the academic side of the institution by cutting non-academic programs and positions. Data from 84 public universities in six states (California, Florida, Georgia, Maryland, South Carolina, and Texas) suggest a more complicated picture. Among the three professional categories, faculty experienced the least growth in positions from 1991 to 1993—.3 percent, compared with 2 and 11.5 percent respectively, for administrators and support professionals. To the extent that there were cuts in positions, they fell on the lowest levels of the job hierarchy—service (-.5 percent) and clerical (-2.5 percent). However, skilled and technical "non-professional" positions increased by .3 and 4.8 percent. Overall, non-professional positions increased somewhat less than faculty (.1 percent), but that slight difference is more than counterbalanced by the continued growth of executive and support professional positions (Montgomery and Lewis, 1995). In the future, scholars need to further explore these patterns in particularly tight fiscal times.

Why do we find such a generalized pattern of growth in non-faculty positions (and thus in non-academic costs), and how do we interpret the pattern? The most generalized accounts of increased administrative costs might be called "a function of" explanations:

> In interpreting statistics to the effect that positions within administrative areas have increased by larger percentages than faculty positions, it should be recognized that such increases are in large measure a function of the increased reporting requirements of government, the growth and complexity of student assistance programs, increased need for counseling accompanying greater nonhomogeneity of the student enrollment, and the increased diversity of institutional goals embraced. (Simpson, 1991, p. 207)

In many such accounts, increased administrative costs are explained as inevitable and unavoidable responses to external stimuli, either external to the institution (e.g., government, demographics) or external to administration (e.g., faculty). Both institutional and resource dependency theory support the view that state regulations and structures can "coerce" organizations to develop corresponding units or "administrative clones" (Catell, 1985) to respond to state demands and to mirror the state (Leslie and Rhoades, 1995). Such explanations are also consistent with the functionalist theory which informs most higher education research. The state is seen as an external regulator to which autonomous organizations (colleges and universities) must respond. And such organizations are seen as operating within the context of various subsystems or environments—political, economic, demographic, technical, cultural—to which they must respond.

In many other such accounts, increased administrative costs are explained as inevitable and unavoidable responses to *internal* characteristics of the organizations. The principal examples of such characteristics are size (Hawley, Boland, and Boland, 1965) and complexity. Generally, organization theorists have argued that complexity is the more significant variable, operationalized in various ways. For example, Blau (1973) focuses on the relationship between differentiation and numbers of administrators, noting "economies of scale"—now part of our common vernacular—in larger universities. Yet, those economies declined over time. Moreover, in reviewing research on economies of scale, Brinkman and Leslie (1986) find a positive relationship between size and complexity, and administrative expenditures. Part of the complexity argument also comes into play in suggesting that certain functions in higher education—e.g., research and sponsored projects—are inherently more administration-intensive (Bowen, 1980; Galambos, 1987). And yet Meyer, Stevenson, and Webster (1985) suggest that the performance of additional tasks is not a principal cause of increased administrative costs.

The problems with "a function of" accounts are at least threefold. First, faculty face increased reporting pressures too. Whether it is in confronting affirmative action requirements (and challenges) or an increasingly diverse student population, faculty are increasingly pressured too. For example, faculty searches, personnel decisions, and the admission of graduate students now require more and more time of faculty (who are often targeted in lawsuits surrounding affirmative action). Similarly, in response to administrative pressures to be more accountable, faculty must spend more and more time developing reports, providing performance indices, and strategically planning. Moreover, in response to an increasingly diverse student population, faculty are under increasing pressure to adapt their teaching styles to various learning styles, to develop more relevant curricula, and to become mentors to larger numbers of students. Why, then, should there not be increased numbers of faculty and of instructional expenditures (expenditures in instructional units)?

A second problem with "a function of" accounts is that state regulations and organizational characteristics do not mandate certain administrative structures. The pressures, external and internal, can be addressed in a variety of ways. "A function of" accounts lack agency. (A major oversight, for as Pondy—1969—found, "administrative intensity" is related to managerial control—administrative costs are less in owner-managed organizations in which "managers" received whatever profits the organization generated.)

The weakness is clear in the case of the most often cited legalistic pressures on colleges and universities (and other organizations)—equal employment opportunity/affirmative action law. The government has not called for the development of specific organizational structures to adequately address personnel matters (Skrentny, 1996). Rather, the translation of federal law and enforcement into organizational policy, structure, offices, and practices has been mediated by various professions in organizations—for example, human relations personnel and labor lawyers (Edelman, Abraham, and Erlanger, 1992; Selznick, 1969; Sutton and Dobbin, 1996). Such professionals aim to establish and/or extend jurisdiction over employment relations. They respond "opportunistically" to developments in employment law.

> In the early 1970s, several articles in the professional personnel literature urged personnel administrators to use managers' uncertainty over standards of compliance as leverage for upgrading and formalizing the personnel function within firms. (Sutton and Dobbin, 1996, p.800)

Such a mediating mechanism sounds very much like the mechanism of "normative isomorphism" in institutional theory (DiMaggio and Powell, 1983). There is an increasing similarity among organizations because of the circulation of similarly trained and positioned professionals, and because of the prevalence of accepted professional norms and practices. Yet Sutton and Dobbin (1996) extend the idea by noting the existence of competing professionals within the same organization—different professionals promote different practices within the same organization.

The implications for higher education are fascinating, and entirely unexplored. The area of employment law potentially impacts the jurisdictions of various professions on campus, from faculty to human relations, to lawyers, to student services, and on and on. Administration consists not simply of "bureaucrats," but of various professionals (what I call elsewhere, "managerial professionals"—see Rhoades, 1996). Thus, we can understand the translation of state regulations and legislation regarding employment into the development of new "administrative" offices and the hiring of new "administrative" personnel law, not as a simple organizational response to external pressure, but as an ongoing, politically negotiated struggle among various occupations and professions on campus. There is virtually no study in higher education of human relations professionals and their position and activity in relation to governmental regulations (for a limited exception see Bouchard, Davidson, and Fortunato, 1992, on the disproportionate growth of human relations staff from 1986 to 1991 in a sample of 49 universities). (The role of professionals in utilizing external environments to the benefit and growth of their profession/occupation is a point that I will take up later.)

A third problem with "a function of" accounts is empirical. There is little support empirically for generalized arguments about external pressures. Such accounts can explain neither the general pattern nor the disaggregated pattern of differential growth for different sectors of administration. For example, one study of human resource personnel staff from 1986 to 1991 in 49 large institutions found that:

> [S]taffing ratios [of human resources staff to full-time faculty] have gone down over this five year period: either human resources departments are getting larger or the institutions surveyed are cutting back staff outside the human resources function. The only deviation from this trend was observed in the area of labor relations. (Bouchard, Davidson, and Fortunato,1992, p.23)

Similarly, a recent case study of administrative costs finds that externally mandated regulations in the areas of affirmative action, environmental health and safety, and collective bargaining account for only a small proportion of the increases in administrative expenditures (Gumport and Pusser, 1995). In that same case, some support was found for the effects of economies of scale, as evidenced in the slightly lower administrator to faculty ratios at the larger than at the smaller University of California campuses. However, the "a function of," functional accounts of increased administrative costs do not discriminate among and between various horizontal, functional divisions of administration and types of non-faculty personnel.

Two other generalist accounts have been advanced, both of which specify mechanisms in the form of managerial actions, that contribute to increased administrative costs. First, Zemsky and Massy (1990, p.20) have coined the term "administrative lattice" to refer to a process of administration:

> endlessly extending itself in response to an environment of regulation and micromanagement, of administration becoming a goal in itself, and of a commitment to consensus management that too often made higher education risk averse.

In directing attention to the consequences of an ideology of consensus management, Zemsky and Massy are referring *not* to shared governance with faculty, but to extensive consultation within central administration.

Related to the lattice is the rapidly "subdividing" and "proliferating" characteristics of administrative staff.

> Colleges and universities have reacted to the increase in the range of their responsibilities by subdividing the duties attached thereto among administrators, each of whom follows suit by again subdividing duties, and so forth, repeatedly.... [W]hat too often happens is that below each vice-president is a linear series of positions, each with secretarial trappings, and each occupied by an individual who conceives of himself or herself as having a policy or at least supervisory role, aloof from performing the substance of the work itself. There is then need for time spent in coordination meetings and for the appointment of someone additional to serve as coordinator. (Simpson, 1991, p.208)

There is a sense of almost mindless, unintentional growth.

A second generalist account is provided by Leslie and Rhoades (1995),with their conception of "organizational distance" and position. Two new dimensions to the discussion are introduced with this concept—differentiated administrative costs by vertical hierarchy, and budgetary authority and power. "Distance" can take on various meanings, from physical separation to vertical layers in the organizational chart (see Glaspar, 1995). The proposition is:

> The greater the organizational distance between the unit and the budgetary decision maker, the smaller will be the proportional increase in the resource allocation to that unit. (p.203)

According to this proposition, we can expect to find differential of rates of administrative growth according to position in the vertical hierarchy. A key point is the locus of budgetary authority in the hands of some administrators. Administrative costs are not simply a matter of external pressures and internal characteristics, of inevitable trends, systemic, disembodied processes, or of managerial belief systems. Instead, the patterns of costs are shaped by the political power and actions of various administrators.

However, it is critical to disaggregate along *horizontal* lines, for administrative costs and growth are not uniform across the board. Among the standardized national expenditure categories of institutional support, academic support, student services, and operations and maintenance, the most dramatic examples are operations and maintenance (O & M), and libraries separated out from the rest of academic support. Both of these divisions have realized patterns quite distinct from those of other administrative/support functions. From 1976/77 to 1989/90, they are the only non-instructional expenditure categories that realized a decline in their share of institutional revenues. That decline was experienced in every sector of higher education. For O & M, and for libraries, respectively, those declining shares were: from 9.1 to 7.8 percent and from 3.5 to 3.1 percent in public universities; from 11.5 to 9.6 percent and from 3.9 to 3.3 percent in public four-year colleges; from 11.2 to 11 percent and from 3.5 to 2.5 percent in public two-year colleges; from 8.8 to 7.5 percent and from 4.2 to 3.4 percent in private universities; and from 11.2 to 9.1 percent (missing data on libraries) in private four-year colleges (Zumeta and Looney, 1994).

In some regards, the pattern in libraries is particularly puzzling. For yet another generalist explanation of increased administrative costs is technological. Many administrative areas draw heavily on technology, which has extensive equipment, training, support, and upgrading. For example, as two observers note in regards to a campus network (including libraries), "Estimates received by campus administrators for 'running the wire' seldom are total costs. The ancillary or support costs can easily triple the original expense." (Gilbert and Green, 1995, p.12) Similarly, information technology must continually be upgraded:

> [T]echnology resources are expensive yet have a short half-life, often less than 15 months. (Gilbert and Green, 1995, p.19)

Technology has extraordinary short-term costs, the benefits of which are very much deferred (Simpson, 1991). In recent years, there may be no other area of support/administration that has been more profoundly impacted by the rapid introduction of new technologies than libraries, from security systems, to computerized check-out systems, to on-line catalogues, to various information technology-based search and data systems. Moreover, libraries are obviously directly connected to the education of students and to scholars' production of research. It is quite puzzling, then, that this support division has experienced such a decline in expenditure shares. The generalist explanations shed little light on the differential experiences of functional divisions of support/administration.

Some accounts have been offered that differentiate horizontally among functional divisions of support/administration. For example, Leslie and Rhoades (1995) articulate propositions that linked differential administrative costs to unit differences in perceived closeness to high technology and corporate marketplaces (see also Slaughter, 1993, for this point in regard to academic units) and in ascribed characteristics of personnel and their clients. The latter point builds on empirically validated findings that a profession's prestige is linked to the prestige of their clientele (Heinz and Laumann, 1982) and that the "feminization" of an academic field is linked to lower salaries, holding other significant variables (e.g., productivity) constant (Bellas, 1994) (the significant independent effect of gender on academic salaries is also empirically validated—Smart, 1991).

Rhoades (1995a) has followed up on these propositions, exploring the "place" of student affairs in the administrative hierarchy and in restructured colleges and universities. Tarred with the brush of "service" in an era in which the prevailing discourse is about "productivity," student services has fared worse among administrative divisions than have institutional support and academic support. Whether the measure is salaries of top administrators or the division's share of institutional expenditures, student affairs is towards the bottom of the administrative hierarchy. It still gained share relative to instruction: in every institutional sector from 1976 to 1990, student affairs increased its share of institutional expenditures relative to instruction. Yet, with the exception of private four-year schools, general administration expenditures represented a larger share of expenditures than student affairs and increased faster than did expenditures for student affairs (Glaspar 1995). Perhaps in response to this pattern, in an effort to become more "productive" and self-sufficient, student affairs has increasingly "privatized" various parts of student services (e.g., residence life) and increasingly begun to charge students fees for services rendered.

In addition to the need to differentiate administrative costs horizontally, it is necessary to move beyond central administration in analyzing such costs. Administrative costs are incurred not just in central administration, but in various other units, such as academic colleges. In calling for caution in analyzing and drawing conclusions about patterns and causes of administrative costs, Simpson (1991, p.207) notes that "The data used [to track administrative costs] may not, for instance, include positions at school and department levels." Similarly, Stanley and Adams (1994, p.128) repeatedly emphasize that there are administrative expenditures in academic units: "Also, neither academic support nor institutional support includes any administrative expenditures of academic departments so that these categories do not reflect administrative expenditures at all levels of the university."

Two case studies provide results along these lines. Stanley and Adams (1994) examine three locations of administrative costs: central administration (offices of president, provost, vice-provost, and vice-presidents); mid-level administration ("units that report to the central administration and whose primary function is coordination of other units"—p.135); and operating units (both academic and non-academic departments). They find that 7 percent of administrator FTEs at one public research university are in central administration, 27 percent are in mid-level administration, and 66 percent are in operating units. (Notably, there is a different distribution of salary costs—9 percent at central level, 31 percent at mid-level, and only 60 percent in the operating units—those findings support Leslie and Rhoades' concept of "organizational distance—those closer to budgetary authority get more revenues.) Perhaps most important, they find that 18 percent of all administrator FTEs are at the operating unit level in academic departments, with substantial variation by departments. Unfortunately, such differences are not indexed against any size or performance measures. Nevertheless, the findings point to a significant amount of administrative costs in academic colleges and departments.

Rhoades (1995b) analyzes administrative costs in instructional units over time. Focusing on nine academic units between 1987-88 and 1991-92, Rhoades finds that increases in administrative expenditures of these units outpaced the growth in central administration's expenditures in most units, both in the expenditure of state monies and of total monies. Moreover, indexing such growth against increases in programs, student credit hours (and FTEs), and faculty FTE reveals that the growth of administrative expenditures in the academic units is out of proportion to increasing complexity or size. The most significant unexplained finding of the study is that there is a great range in the pattern of increases from one academic unit to another. In other words, not only are administrative costs in academic operating units significant, there are significant variations in such costs among these units.

Such college level administrative costs may be especially important in particular administrative/support areas—e.g., fund raising. For example, Grunig (1995) analyzes the structure of development offices, focusing on the decentralization of positions and activities to the colleges. Many colleges, especially professional schools, have their own development officers. Grunig (1995, p.697) concludes that:

> The results of this study suggest that factors other than the pursuit of fund-raising efficiency may be principally responsible for the trend toward greater decentralization of development offices....Amid the current climate of funding cutbacks it is difficult for many institutions to justify significantly increasing development budgets while budgets for other purposes are being reduced or eliminated....However, because the salaries of most constituent development personnel are either partially or fully paid by the college

> units they represent, decentralization allows new development personnel to be added while keeping the budget increases for the main development office small. In many cases, decentralization appears to be used as a form of organizational growth, rather than solely as a form of reorganization.

It is critical that decentralization is correlated with *more* development personnel and higher expenditures, but *not* with greater fund raising efficiency, measured as "cost per dollar raised."

Of course, administrative costs are incurred not just at the college, but at the departmental, or operating unit, level. No doubt, some of these costs attach to units' production of teaching and research, whether in the form of administrative assistants, secretaries, advisors, graduate and undergraduate program coordinators, grants writers, or research technicians in laboratories. There is virtually no research on such personnel, many of whom are included in aggregated data as "instructional expenditures"—that is, expenditures incurred in instructional units. There is, then, an extraordinary *under*estimation of administrative costs.

Some such administrative costs in departments (and colleges) may be incurred due to central administrators' actions to shift responsibilities and/or costs from central administration to operating units. In a time when the costs of central administration have come under increased scrutiny, such accounting and responsibility shifts may be undertaken to deflect criticism and demonstrate contained costs at the central level.

In some types of institutions, there are considerable administrative costs in the operating units that attach to research. Some of these costs are incurred in traditional, discipline-based departments. The costs may include capital (equipment), operations, and personnel (e.g., technicians in labs) expenditures. In addition, some administrative costs are incurred in the growing number of interdisciplinary centers and institutes—so-called organized research units (Geiger, 1990). A recent survey of such units in the fastest growing (in the research rankings) research universities indicates that such centers account, on average, for 28 percent of institutions' research expenditures (Stahler and Tash, 1994). They are a major part of research production on campuses. More important for my purposes here, such organized research units tend to have more hierarchical and well developed administrative structures than do departments (Friedman and Friedman, 1984; Hays, 1991; Teich, 1982). In part, this has to do with the interdisciplinary coordination and grants oriented activities of the units. In addition, it has to do with the efforts of such units to develop, maintain, and oversee ongoing relations with private industry.[1] Partnerships between universities and business have not only been initiated by institutions seeking to enhance relations with the private sector and to generate new revenues, they have been encouraged by federal

---

[1]My thanks to Brian Pusser for raising this point after a presentation of mine on longitudinal change in federal science and technology policy (Rhoades and Slaughter, 1994). Pusser asked whether the increased emphasis on centers that interacted with industry led to increased administrative costs.

science and technology policy (Rhoades and Slaughter, 1994). Engineering Research Centers are a prime example of such units, sponsored and underwritten by the National Science Foundation (Mayfield and Schultzman, 1987). The ongoing cultivation of such external relations takes personnel, resources, and time.

Much of the entrepreneurial activity surrounding research originates at the central administrative level. In a subsequent section I trace the emergence of one such occupational grouping of professional and support employees to highlight one of the drivers of "administrative growth." For now, suffice it to say that central administration certainly has invested in a variety of ways in what has been "tactical administrative growth"—the growth of units that are designed to generate revenues (Gumport and Pusser, 1995).

In going beyond central administration in studying administrative costs, it is necessary to look not only below, but above, the center. There are considerable administrative costs, for example, in state boards.

> The fastest growing segment [in administrative costs] in percentage terms in higher education budgets in Illinois is the Illinois Board of Higher Education....Other states committed to elaborating programs of assessment, greater coordination, or more intensive scrutiny have also seen a growth in agency administrative and professional personnel. Such agencies tend to be classic public bureaucracies. They grow by adding personnel to undertake more analyses and more evaluations, and implicitly encourage legislators to seek more data and greater "accountability." (Andersen, 1991, pp. 20-21)

This from a former deputy vice-chancellor for academic affairs. The suggested shift in the role of state boards is significant. At one time, boards were seen, and saw themselves, as promoters and defenders of higher education. In recent years, public colleges and universities have come to be not so much buffered from external pressures by these bodies as buffeted by them.

It is not just boards that sit above individual campuses. Other layers of administration rest in the central offices of multicampus districts, most often found in community colleges. Here, Glaspar (1995) has found a significant site of administrative growth, in numbers, salaries, and in capital equipment and building expenditures. In some state systems, such as California, there is an additional layer of systemwide administration. The case of California may represent the extreme in this regard, with significant sized offices for the University of California, the California State University System, and for Community Colleges. The growth that one finds at the central level of individual campuses is often matched by growth at the systemwide level (Gumport and Pusser, 1994).

It is essential, then, to disaggregate administrative costs. The literature has not moved very far along these lines. However, the little work that exists is suggestive of some important directions for future research.

As noted in introducing this chapter, it is also essential to rethink and reconceptualize our conceptions of administrators and of costs. The framing of administrators versus faculty does not capture the complexity of these employees' lives. In the ensuing section, I utilize an empirical case to dramatize the limitations of existing classification categories. Many employees simply do not fit in a dichotomous frame. Such binary categories also fail to capture the restructuring of professional employees in higher education, overlooking the "production" role increasingly played by non-faculty personnel. Thus, in a later section I provide another empirical case of non-faculty employees to dramatize the need to consider the production work and thus the "productivity" of non-faculty personnel, studying them not simply in terms of "costs," but in terms of "investments" and "yields."

Empirical Cases: Unionized Non-Faculty Personnel

The fallacy of classifying all non-faculty personnel as administrative "other" is dramatized by unionized support staff, academic professionals, and non-academic professional and administrative personnel. Such personnel do not fit neatly into the dichotomous and adversarial schemata of faculty versus management. They violate the schemata in a fundamental way. They are not faculty. But neither are they management. In fact, such unionized personnel have organized collectively to negotiate conditions and terms of work with "management." Faculty is "other." So is management.

Who are these non-faculty, non-management personnel? The personnel who are included in bargaining units vary between two general categories—Educational Support Personnel and Academic Professionals. They also vary within these two categories, from contract to contract.

The categories, Educational Support Personnel and Academic Professional, are utilized in many analyses of higher education personnel, and in a software package of contracts collected by the National Education Association and put on CD-ROM—the Higher Education Contract Analysis System (HECAS). These collective bargaining agreements include contracts negotiated not just by higher education unions such as the National Education Association (NEA) and the American Federation of Teachers (AFT), but also by major unions of service workers, such as the American Federation of State, County, and Municipal Employees (AFSCME) and the Service Employees International Union (SEIU) and by other bargaining agents as well.

Nationwide, there are almost 50 different national unions representing some groups on campus. However, AFSCME and SEIU are the two dominant unions representing non-faculty staff, accounting for 44.7 and 23% respectively of unionized staff on campuses. There are 373,216 such workers represented by bargaining units on 881 campuses: 42.8% (131,232) of blue-collar workers, 37.2% (153,462) of clerical workers, and 14.8% (97,111) of professional and technical employees (Hurd, 1995). As with faculty (see Rhoades, 1993; forthcoming), most unionized non-faculty staff are in the public sector: of the 2,722 postsecondary institutions with 500 or more students, public sector campuses account for 84.6% of unionized employees (Hurd, 1995).[2]

The personnel within these various categories share in common the fact that they negotiate collectively with management. Yet, in some sense they are in competition with each other and with faculty who negotiate collectively with management as well. For example, some faculty contracts include clauses that restrict the numbers and ratio of non-faculty personnel, identified as "administrative." That includes support personnel, and in some cases, academic professionals. A union, such as the NEA, which is the representative for several such categories of personnel, is in some sense limited in what they can say about patterns in the numbers of faculty versus non-faculty personnel. In the case of academic professionals, many faculty contracts include these personnel; many others do not. Indeed, the history and ongoing story of academic professionals is of professional personnel seeking the same rights and privileges as faculty.

To illustrate the variations within and between the categories of Educational Support Personnel (ESP) and Academic Professionals (AP) consider the following contracts. Most of the ESP contracts cover primarily classified workers who fall within the general EEOC category of "non-professional" personnel. Yet even here there is much variation, among technical and paraprofessional, clerical and secretarial, skilled crafts, and service and maintenance personnel, with professional and supervisory personnel sometimes included. For example, at Central Michigan University 98 job titles (and 137 members) are included just in the NEA's ESP bargaining unit (and there are five other unions representing 728 other maintenance, clerical, maintenance and food service, police, and clerical workers on the campus).[3] Some are clerical workers who might be found in central administration as well as in the academic units—Administrative Aide I & II. Some are technical workers in science units who are associated with the production of research—Biology Lab Technician; Chemistry Lab Technicians/Stockroom Managers I & II; Electronics Microscope Facility Supervisor; Physics Lab Technicians I & II (science lab technicians usually have Masters degrees); Piano-Organ Technician. Some are technical and/or paraprofessional personnel in various student services—Admissions Specialist; Financial Aid Record Specialist, Financial Aid Specialists I & II. Some are craft and supervisory personnel in the maintenance and facilities branch of central administration: Maintenance and Repair Technicians I & II; Supervisor, Building Services; Supervisor, Carpentry Services; Supervisor, Landscape Operations; Supervisor, Maintenance Mechanics; Supervisor, Powerhouse. Some are technical and/or paraprofessional personnel in academic support services in the library: Library Acquisition Specialist; Library Bibliography Specialist; Library Circulation Specialist; Library Government Document Specialist; Library

---

[2]In one sense, the numbers are deceiving, for public sector units tend to be larger in numbers of staff covered. Focusing on campuses, 47.9% of the public and 17.3% of private sector campuses have at least one staff bargaining agent.

[3]Those unions are, respectively, Communications Workers Association (CWA), United Auto workers (UAW), AFSCME, an independent bargaining agent, and another local of the UAW.

Interlibrary Loan Specialist; Library Reserve Specialist. Some are technical and/or paraprofessional personnel in various business services, some perhaps located in central administration and some located in the academic colleges: Accounting System Technician; Accounting Systems Specialist I & II; Payroll Specialist I & II; Payroll Technician; Personnel Specialist; Receivable Accounts Specialist. Some are clerical and technical personnel connected with computer and telecommunications services and operations, in various sectors of administration: Computer Operator I, II, & III; (Senior) Computer Repair Technician; Switchboard Operator; Telecommunications Repairer/Installer I &II. And all that still does not cover 54 job titles. (Note that some titles include the term, "supervisor." However, they are apparently such low level supervisors in the managerial hierarchy that they do not qualify as a "manager, executive, administrator." Similarly, many "technicians" have college degrees, including Masters' degrees, but again they are apparently such low level professionals that they are classified as technical rather than as professional personnel.)

By contrast, the ESP contract at Youngstown State University is somewhat different in the types of personnel included in the bargaining unit (NEA). It covers 107 job titles (and 400 employees), and it is more inclusive than the Central Michigan contract in that it encompasses some 15 "permanent part-time" position classifications (in addition, there is another NEA local covering 100 administrators, and an independent agent representing 20 police officers). There are five types of laboratory positions—Animal Technician 1, Machinist 1, Technician I & II, Technologist—but none are specified to a particular department. There are several "supervisor" titles, but more than the largely maintenance and trades focus of Central Michigan's positions: Building Maintenance Supervisor I & II; Computer Operations Supervisor; Custodial Work Supervisor; Records Management Supervisor; Storekeeper Supervisor. There are more computer-related job titles: Computer Operator 2; Programmer Analyst 4 & 5; Software Specialist; Systems Analyst 1, 2, & 3. There is also a Plant Maintenance Engineer. In addition, there are similar positions in student services, business services, and operations and maintenance, although some of the titles are different: Account Clerk 2 & 3; Accountant 2 & 3; Assistant Auto Mechanic; Automotive Mechanic 2 & 3; Carpenter; Custodial Worker; Employee Benefits Coordinator; Maintenance Repair Worker 1, 2, & 3; Plumber; Purchasing Assistant 1 & 2; Student Loan Specialist; Student Services Counselor. All of the above are regular full-time personnel. In addition, some permanent part-time personnel are included in the bargaining unit. All but two of these classifications are in central administrative units (the two exceptions are Data Entry Operator 2, in the Cushwa Center for Industrial Development, and Laundry Worker in Human Performance and Exercise Science). These positions tend to be more entry level sorts of jobs: Data entry Operator 2 & 3; Technical Typist; Secretary; Sales Clerk; Clerk; Account Clerk. Again, the variety is enormous.

Somewhat more restrictive in character is the bargaining unit at Southern Illinois University at Carbondale (NEA). There are 91 classifications (and 650 employees), yet 20 of these are various types of operators (e.g., I, II, & III Accounting Machine; I, II, & III Data Entry; I, II, & III Digital Computer; I, II, & III Word Processing), 7 are Secretaries or Office System Assistants, and 11 are Clerks (e.g., I, II, and II Clerk, Accounting; Clerk, Inventory; Clerk, Staff; I, II, & III Clerk, Typing). (There are also 9 other unions representing 547 laundry and custodial, electrical technical, drivers, maintenance, engineers, printing, grounds and food service, police, and trades personnel.) As at Central Michigan, there are some personnel that are clearly housed in the academic units, involved in research: Assistant I, II, & III, Natural Science Lab; Assistant, Natural Science, Technical; Assistant, Physical Sciences Staff; Assistant, Physical Sciences, Technical; Assistant, Research Engineering; Supervisor, Research Lab Shop; Technician I & II, Language Lab. There are also some positions related to research at the central administrative level: Caretaker, Laboratory Animal; Technician I & II, Laboratory Animal Care. And, of course, there are various positions housed within student services, business affairs, and operations and maintenance in central administration: Accountant I & II; Advisor I, II, & III, Financial Aid; Account Technician I & II; Coordinator of Parking and Traffic; Draftsman I & II, Architectural; Draftsman I & II, Engineering; Technician, Electronics; Worker III, Building Service.

Northern Montana College's bargaining unit (AFSCME), partly perhaps because of the size of the institution, is especially narrow in the extent and type of positions. There are 46 classifications (and 80 employees) included in the unit. These consist largely of service and clerical positions, with few technical and only two supervisory positions (Custodial Supervisor I; Groundskeeping Supervisor III), housed largely in central administration. For example, the classifications include: Administrative Assistant I & IV; Admissions Evaluator I; Clerk Accounting, I & II; Clerk, File IV; Clerk, Typist II & III; Cook I, II, & III; Custodial Worker III & IV; Duplicating Machine Operator I; Electrician, Journeyman; Equipment Mechanic III; Food Service Worker I & II; General Office Clerk II; Groundskeeper III; Kitchen Bakery Helper I; Maintenance Worker I & II; Receptionist II; Secretary I, II, & III; Security Guard III; Switchboard Operator III.

The scope of Northern Montana's bargaining unit in terms of hierarchical level of the job positions included in the bargaining unit is comparable to some of the units in community colleges, which tend to be more restricted in that regard. However, there are some interesting institutional type-based differences that emerge in looking at ESP units in community colleges. In particular, there are classifications that fall under "administration" at the aggregated level, but that at the disaggregated, operational level are actually jobs that involve delivering instruction.

An example of a very restricted ESP unit is found in the contract of the Montana Vocational Tech Center. Only 26 positions are included in the unit. One has a supervisor label—Retail Sales Supervisor. Another 6 are labeled technicians: Accounting Technician I & II; Computer Operator Technician I; Library Technician III; Payroll Technician; and Purchasing Supply Technician. The other classifications are clerical, with the exception of two positions that could be coded as instructional—Interpreter/Tutor and Teacher Aide III.

Southwestern Oregon Community College's ESP unit (AFT) includes 79 positions (and 91 employees). However, 11 of these are secretary positions in various administrative units: Admissions; Associate Deans; Business Development Center; Student Support Services; Government Contracts. There is one Supervisor (Equipment/Supplies), one Director (Child Care Center), two Assistant Directors (Financial Aid, Success Center) and four Coordinators (CCLS Outreach, Information Services, and User Services I & II). Again, most of the positions are clerical, in various branches of central administration: Accounts Payable Bookkeeper; Accounts Receivable Bookkeeper; Admissions Clerk; Business Development Center Assistant; Career Center Clerk; Cashier; Computer Operator I & II; Custodian; Financial Aid I, II, & III; Groundskeeper; Maintenance I; Security Guard I & II; Switchboard Operator. There are a few, more skilled, positions: Graphic Artist; Senior Programmer/Analyst; and various library technicians. However, what is most distinctive about the unit is that it includes 9 Teacher Aides: ABE/GED A & C; Automotive; Computer Lab; Curry County; ESL; Life Skills; Music; and Workforce 2000. The distinction, then, between teaching and "administration" breaks down. People in these ESP positions, whose salaries would normally be calculated as part of administrative costs, are directly involved in delivering instruction.

Such overlap between "administration" and "instruction" is particularly clear in the contract of the San Diego Community College District, whose bargaining unit is the "Office-Technical Unit" Chapter #714 of the California School Employees Association (there are two other bargaining units covering 306 workers). Of the 90 job classifications included in the bargaining unit (covering 549 employees), 30 are Instructional Assistants and 12 are Instructional Lab Technicians, housed in academic units. A few other titles include technicians, specialists, a graphic artist and two sign language interpreters. The remaining classifications are clerical: Account Clerk; Administrative Secretary; Bookstore Sales Clerk; Bookstore Sales Clerk, Assistant; Clerical Assistant; Clerical Assistant, Senior; Desktop Publishing Clerk; Inventory Clerk; Offset Press Operator; Secretary; Student Services Assistant; Student Services Assistant, Senior; Word Processing Operator.

At the extreme of inclusiveness among ESP units is the Florida State University System contract, with four certification units of AFSCME. The Florida case is especially important in that it is so large (12,886 employees, with two other bargaining units covering an additional 390 employees). Four categories of employees are included in the bargaining units: Administrative and Clerical, Operational Services, Human Services, and Professional.

> The contract defines 52 Administrative/Clerical position types as: All full-time and part-time University Support Personnel Systems employees whose work primarily

> involves the production, filing, distribution and/or examination of documents or records. Employees in this unit share similar working conditions, most notably an office environment; work does not generally involve heavy physical labor, and similar equipment is used, such as typewriters, telephones, and other office equipment. (Appendix A, p.44)

Specific titles include: Clerical Aide; Clerk; Senior Clerk; Word Processing Operator; Senior Word Processing Operator; Secretary; Senior Secretary; Office Assistant; Office Manager; Purchasing Agent; Senior Purchasing Agent; Fiscal Assistant; Senior Fiscal Assistant; Financial Assistance Counselor; Editorial Assistant; Library Technical Assistant; Senior Security Guard. In addition, there are several positions that seem to be more technical and even professional, highlighting the inadequacy of existing job classifications: Computer Support Specialist; Computer Programmer; Computer Programmer Analyst; Senior Computer Programmer Analyst; Photographer; Scientific Photographer; Cinematographer/Videographer; Art/Publications Production Specialist; Senior Art/Publications Production Specialist; Assistant Editor; University Union Program Specialist; Medical Records Specialist; Police Service Technician.

> The contract also specifies 85 Operational Services positions: All full-time and part-time University Support Personnel System laborers, technicians, mechanics, operators, and service workers whose work involves fabrication, maintenance, and repair activities and/or the provision of personal and domestic services. Work is often performed outdoors and frequently involves heavy physical labor. The basic education and training requirement is graduation from high school, with practical experience in the specific area of work, although some of the skilled workers or technicians may be required to have some type of advanced technical or vocational training. (Appendix A, p.41)

Specific titles include: Computer Repair Technician; Senior Computer Repair Technician; Broadcast Specialist; Television Camera Operator; Teaching Laboratory Specialist; Broadcast Engineering Technician; Engineering Assistant Designer; Engineering Technician; Laboratory Technician; Senior Laboratory Technician; Chef; Catering Specialist; Food Service Worker; Laundry Service Specialist; Launderer; Motor Vehicle Operator; Senior Motor Vehicle Operator; Heavy Equipment Operator; Senior Heavy Equipment Operator; Laborer; Carpenter; Roofer; Cabinet Maker; Plumber; Mason; Steamfitter; Welder; Refrigeration Mechanic; Senior Refrigeration Mechanic; Piano Technician; Instrument Maker-Designer; Scientific Glassblower; Machinist; Laboratory Machinist Specialist; Custodial Worker; Senior Custodial Worker; Custodial Worker Trainer; Animal Technician Aide; Animal Technician; Security Guard; Campus Parking Patroller; Radiation Control Technician; Building Safety Inspector.

> The contract further defines 25 Human Services positions: All full-time and part-time University Support Personnel System employees providing direct human service care, treatment, and rehabilitation to clients, patients, and/or students. Work is often performed in an institutional setting or an office or clinic. (Appendix A, p.43)

Specific titles include: Child Care Group Leader; Teacher's Aide; Health Support Aide; Health Support Technician; Senior Health Support Technician; Certified Radiologic Technologist; Therapy Aide; Respiratory Care Technician; Licensed Practical Nurse; Senior Licensed Practical Nurse; Medical Technologist; Senior Medical Technologist; Dental Assistant; Dental Hygienist; Dental Technician; Senior Dental Technician; Waiter/Waitress; Host/Hostess.

> Finally, the contract defines 67 Other Professional classifications: All full time and part-time University Support Personnel System employees occupying positions in classifications which meet the requirements of a "professional employee" as set forth in Section 447.203(13), Florida Statutes (1985). Work is predominantly intellectual and varied, rather than routine and manual, and involves the constant exercise of discretion and judgment. The training and experience requirements for professional employees typically require that they must have pursued a course of study in a particular field and have acquired advanced knowledge in that field, and their job duties typically require that they be capable of applying this professional expertise in the course of performing their work. (Appendix A, p.44)

Many of the specific titles include positions that most likely fall within the purview of central administration, in various support units: Collections/Loans Manager; Telecommunications Specialist; Telecommunications Technician; Senior Telecommunications Specialist; Computer Support Analyst; Systems Programmer; Senior Systems Programmer; Archivist; Statistician; Senior Statistician; Graphic Artist; Medical/Biological Illustrator; Marketing Specialist; Landscape Designer; Engineer; Senior Engineer. Other positions fall within either specific academic units or within the scope of research activities: Electron Microscope Technician; Senior Electron Microscope Technician; Electron Microscope Manager; Biological Scientist; Senior Biological Scientist; Chemist; Senior Chemist; Senior Teaching Laboratory Specialist; Geologist; Research Physicist; Senior Research Physicist; Health Physicist; Senior Health Physicist; Psychological Aide; Animal Anesthetist. The latter are non-faculty positions, but they are positions—and expenditures incurred—within academic and/or research units. We can begin to get a sense of how vastly underestimated are administrative and support costs when we focus only on the central administrative level. We can also begin to get a sense of how much in the way of administrative and support costs are incurred in pursuing one of the basic missions of universities—research. Such costs (investments) are rarely factored into our one-sided calculations of how much in the way of external grants and contract monies are brought into the institution (yields).

There is a tremendous range and variety, then, both horizontally and vertically, of non-faculty positions in colleges and universities. The work force goes far beyond the faculty. The non-faculty work force goes far beyond "administration" or "management." And this is not even including academic professionals, whom I turn to now.

In most cases, unionized academic professionals are included within faculty bargaining units. In the HECAS data base, nearly half (99) of the faculty collective bargaining agreements include academic professionals. Just 17 contracts cover only academic professionals. And many of these units include administrative personnel. A few examples suffice to make my point: it is misleading to treat non-faculty personnel as residual "administrators." Some "administrators," and academic professionals—who may be included as part of "administrative costs"—have organized collectively to bargain as an adversary with "management."

As with ESP employees, there is a great deal of variety in the types of personnel covered in AP contracts. At Youngstown State University there is an AP contract for the Professional/Administrative Staff Association, including 70 full-time and 32 part-time positions. The major break hierarchically is between the 33 "Coordinator" and 2 Assistant Director" positions, which are included, and the "Director" positions, which are excluded from the unit. The diversity of positions, in terms of the different functional areas of the university, and the location in academic as well as in central and support units, is striking. Consider, for example, the differential locations of the coordinators: of Athletic Business Operations; of Career Services; of Continuing Education; of the Curriculum Resource Center (in Education); of Early Advisement and Registration; of Intramural Programs; of the Language Laboratory (Foreign Languages); of Learning Services (Nursing); of the MBA Program; of the Mathematics Laboratory; of Programs (Urban Studies); of the Reading Laboratory; of Special Student Services; of Student Grants; of Student Loans; of Testing; of Undergraduate Advisement; of the Writing Laboratory. The diverse functional location also applies to the other positions, which include: Academic Advisor; Administrative Aide; Adult Admissions Counselor; Administrative Assistant (Deans); Assistant Acquisitions Librarian; Assistant Director of Admissions; Assistant Director of Scholarships and Financial Aid; Assistant Reference Librarian; Athletic Trainer; Catalog Librarian; Counseling Psychologist; Database Administrator; Equipment Manager; Fine Arts and News Announcer/Producer; Grants and Contracts Administrator; Housing Coordinator; Planetarium Lecturer; Research Coordinator (Graduate School).

Similar variety is evident in several technical and community college contracts, with an interesting variation on the hierarchical theme. The Maine Technical College System's AP contract, for example, is with the Administrative Unit of the Maine Teachers Association. It includes 68 positions at five levels of administration. Directors and Associate Deans are included in the bargaining unit, which is concentrated overwhelmingly on central administrative positions. Included are the following positions: Financial Aid Director; Director of Development; Director of Admissions; Associate Dean of Academic Affairs; Associate Dean of Students; Assistant Dean of Continuing Education; Assistant Dean of Nursing; Assistant Dean of Instruction; Assistant Dean of Students; Tech Prep Director; Director, Curriculum Resource Center; Coordinator, Business and Industry; Coordinator, Financial Aid; Associate Director of Admissions; Automated Systems Manager; Assistant Registrar; Assistant Tech Prep Coordinator; Grants

Administrator; Conference Coordinator; Coordinator, Financial Aid; Coordinator, Truck Driving; Grant Developer; Day Care Director; Success Counselor.

At Macomb Community College the AP contract, negotiated by the Association of Administrative Personnel, includes 44 administrative positions, which run up to the level of Associate Vice-President for Academic Affairs. Such a pattern raises further questions about what drives administrative costs in some institutions, even at the central administrative level. At one and the same time, Macomb's contract points to the proliferation of administrative units at the central level, paralleling administrative positions in academic units, and it calls attention to administrative costs in academic units, normally not included in discussions of administrative costs. Consider the Deans and Associate Deans, found both in central administration and in academic units. There are Deans of Academic and Student Development Services; Arts and Sciences; Business, Health and Public Service; Student and Community Service; and Technology. Included among the 19 Associate Deans are those of Admissions and Academic Services; Arts and Sciences (one each for English Language and Literature, Humanities, Mathematics, Physical Education, Science, Social Science); Business and Public Service (one each for Business Administration, Information Management, and Public Service); Cooperative Education and Employment and Transfer Services; Counseling and Special Services; Financial Aid, Resource and Information Management; Health and Human Services; Technology (one each for Applied Technology, Design Technology, and Mechanical Technology). Thirteen of the 19 associate deans (and all 4 Assistants to the Dean) are in academic units. By contrast, 8 of the 10 Director positions are in non-academic units.

Instead of listing in full the positions included in the bargaining unit, some contracts identify descriptors of included and excluded personnel in the "Recognition" clause of the contract. For example:

> The Board recognizes the [Oregon Education] Association as the sole and exclusive representative for all full-time and regular part-time (one-half time or more) annually contracted employees. This includes instructors, librarians, counselors, nurses, coaches, and department chairpersons, and excludes supervisors who effectively make recommendations on hiring, promotion, and assignment. (Clackamas Community College, Article 1)

Other contracts refer to state rulings and/or laws regarding membership.

> The Board of Governors hereby recognizes the Community College of Rhode Island Professional Staff Association/National Education Association of Rhode Island/ National Education Association as the sole and exclusive bargaining agent for all collective negotiations set forth in the General Laws of Rhode Island Title 36-11 for all employees of the Community Colleges of Rhode Island who are members of the bargaining unit outlined in the Rhode Island Labor Relations Case No. EE326B. (Community Colleges of Rhode Island, Article 1.1)

> Pursuant to and in accordance with all applicable provisions of Act 379 of the Public Acts of 1965, as amended, the Employer does hereby recognize the Michigan State

> University Administrative–Professional Association as the exclusive collective bargaining representative for those employees in the defined bargaining unit for the purpose of collective bargaining with respect to rates of pay, wages, hours of employment, and other conditions of employment. On August 2, 1985, the Michigan State University Administrative–Professional Association was certified by the Michigan Employment Relations Commission in Case No.R85 H-207 as the representative for all regular full-time and regular part-time Administrative and Professional employees serving in a non-supervisory capacity and located upon the main campus of Michigan State University....Excluded are: Executive, Supervisors, and Confidential Employees. (Michigan State University, Article 1.1 & 1.2)

> The Board hereby recognizes the Association as the sole and exclusive bargaining agent for all employees within the bargaining unit. The bargaining unit consists of those classes of positions found appropriate as a result of the petition submitted in case number EE-2071 and any such positions which may be added or deleted by the State Labor Relations Board. (University of Rhode Island, Article I)

Thus, the state is more than simply a regulator of higher education, in the sense of issuing regulations that lead to increased administrative costs. Its various arms also define categories of workers and determine whether these workers can organize collectively. The bases of such determinations are in part the extent to which the employees in question have supervisory and managerial responsibilities. There are other criteria and factors as well, as evidenced in decisions regarding the organization of part-time faculty (Rhoades, 1996).

The details as to what employees are included in bargaining units reveal that we must disaggregate "administrative other" in our analyses of administrative costs. Our current categories are inadequate. They do not capture the complexity of the non-faculty work force.

More than that, the collective bargaining agreements point to the role of agency in shaping administrative costs. The growth and costs of various non-faculty groups on campus is a matter not of "bureaucratic accretion," or "organizational distance," or of an impersonal "administrative lattice." Rather, these matters are the subject and the outcome of contract negotiation.

Empirical Cases: Technology Transfer Professionals/Administrators

The fallacy of classifying all non-faculty personnel as administrative "other" is dramatized not just by examining such staff who are unionized. It is also highlighted by examining non-faculty personnel who are directly involved in "production work" in higher education. I focus here on technology transfer professionals and their staff. In exploring the growth and activities of such professionals, I offer a case in point of why we must reconceptualize non-faculty personnel and the causes of their growth, and why in the process we must attend more to the "productivity" of such employees.

In the previous section on unionized non-faculty personnel, it was clear that some of these employees are production workers. For example, teaching aides and research lab technicians do not simply support the two central missions of colleges and universities—teaching and research. They actually conduct teaching and research. They are part of the production process. Thus, just as we can gauge faculty's "productivity" in these areas, so too we can gauge the productivity of such non-faculty personnel. Indeed, we already have begun to think about graduate teaching assistants in this regard, determining the proportion of student credit hours they generate, particularly at the lower division undergraduate level. It is more challenging to think about how one would gauge the productivity of a teacher's aide, but there must be some student outcomes that could be correlated with the presence (or absence) of teacher aides. Similarly, although I am unaware of any efforts to calculate the "productivity" of various lab workers, there is no reason to believe that one could not generate a range of indices according to which one could measure lab technicians' contribution to the research effort.

As soon as one starts moving in this direction, it makes sense to start thinking about non-faculty personnel not simply in terms of "costs," but in terms of "investments" and "yields." We need to become better accountants in higher education, recognizing that in utilizing many non-faculty personnel, we are not only incurring increased costs, but we are investing in increased yields. Whether the investment is "paying off" in either a short- or a long-term sense is an empirical question we as a field have yet to pursue.

The issue of productivity, and the question of whether our increased investments are resulting in increased yields, is even clearer in the case of non-faculty personnel whose responsibility—direct or indirect—is to generate revenues. The mission that technology transfer professionals are fulfilling is service. The traditional conceptualization of public service, perhaps best epitomized by extension services in agriculture, has been transformed to a conception that involves generating revenue for the institution at the same time the institution contributes to the economic development of the state. The shift is a conscious and explicit one, articulated by university leaders (Slaughter and Rhoades, 1993). Technology transfer professionals are presented as producing service for the community, which depending upon one's position means creating more marketable products, jobs, and companies, and contributing to the state's tax base (Rhoades and Slaughter, 1991). Yet such professionals are also presented as providing a service for the institution—they increase the commercially relevant research activity of the faculty and assist in the commercialization of faculty's intellectual property in ways that brings increased revenues for the university.

Technology transfer professionals have only emerged in higher education within the past quarter century. In 1974 the Association of University Technology Transfer Managers (AUTM) was created. Its first professional staff was hired in 1986, when the membership of the association was 381. By 1990 it had 709 members. Its membership currently numbers about 1,600.

Of particular note is the continued growth of these professionals in a period of fiscal stress in higher education and in a time when institutions claim to be targeting non-faculty personnel for reductions more than faculty. From FYs 1992 to 1994, AUTM's national survey of members reveals that the number of professional FTEs for licensing activities in universities grew from 196.6 to 238.82, an increase of 21.5 percent. The number of professional FTEs for technology transfer grew from 288.45 to 356.85, an increase of 24 percent. Obviously, the absolute numbers are far from overwhelming. Yet, such growth in a two-year period in such a relatively small area of activity is striking.

Additionally, it is striking how widespread such growth is in terms of numbers and rankings of institutions. There is much growth in such personnel not just in those relatively few elite institutions that generate a lot of patenting and licensing activities and royalties, but also in a range of institutions that generate very little in the way of royalties. Consider the growth of FTEs for technology and for licensing activities in the second 50 institutions (ranked by numbers of professional FTEs). From 1992-93 professional FTEs for technology transfer and for licensing grew 11.65 and 44.01 percent respectively. Support FTEs for technology transfer and for licensing grew 36.44 and 48.22 percent respectively. From 1993-94 there were slight decreases in professional FTEs (2.02 and 5.33 percent respectively), yet continued increases in support FTEs (11.91 and 22.55 percent respectively). Consider the gross royalties received in 1994 in many of these institutions: $0—San Diego State University, Illinois State University; $7,300—University of South Carolina; $10,000—Northern Illinois University; $12,000—Ohio University; $14,326—University of North Carolina Charlotte; $62,000—Drexel University; $65,117—Montana State University; $67,762—Ball State University; $102,350—Syracuse University; $110,000—Brandeis University; $112,435—University of Akron; $153,777—Georgetown University; $169,119—Dartmouth College; $173,750—University of Central Florida. Many other "second 50" institutions are making considerably more in royalties than the institutions I have identified. My point is simply that there are a large number of institutions investing in technology transfer and licensing personnel even in the absence of considerable royalty revenues.

To the extent that we reconceptualize such non-faculty professionals as production workers, we can begin to ask questions about their productivity. In the case of employees such as technology transfer professionals, that productivity can be measured in various objective ways, including invention disclosures received, patent applications filed, patents issued, and licenses and options executed. However, as noted earlier, the principal rationales for investing in such technology transfer activity is the generation of revenues for the institution and wealth (or jobs) for the state. Thus, such "bottom line" measures are probably also reasonable criteria for gauging performance. Obviously, consideration has to be given to the time it takes to bring products (and businesses) to the marketplace and to the point where they are generating revenue. But at the very least, some time frame and/or time line should be utilized in guiding decisions about how much (and whether) to invest in certain activities given certain yields.

Existing data can only take us so far in conducting such cost/benefit analyses. For example, AUTM's survey provides data only on numbers of FTEs, not on the expenditures or even the budgets of the relevant offices. Moreover, although there is data on "legal fees," these refer to the routine fees incurred in prosecuting patents and the like. Such figures do not include extraordinary costs incurred (in attorney fees and out-of-court settlements) in patent litigation—for example, surrounding patent infringement. At my own institution, the university recently agreed to an out-of-court settlement of $1.5 million in a suit filed against several faculty and officials of the university. That more than counterbalances all the royalties that have been collected in the past 10 years. If we are to truly gauge the "payoff" or the "payout" that attaches to university involvement in technology transfer activities, we need to begin to gather this and other data about a broad range of efforts.

Just in the area of technology transfer, to focus only on AUTM is obviously to vastly underestimate universities' investment in technology transfer. For example, a recent national survey of various technology transfer units on campus identified several types of units, including small business development centers, research and technology centers, incubators, and investment units (Dill, 1995). Of these, offices of licensing and patenting tend to be the smallest in terms of FTEs—an average of 2.9 professionals, and 2.3 support staff, with a range from 1-20, as compared with 9 and 14.6 respectively for research and technology centers, and 50.5 and .75 for incubator facilities. Even so, the average budget of licensing and patenting offices is $307,216, with a range from $20,000 to $2 million. So this is not an insignificant area of university investment. Such entrepreneurial efforts require managerial and technical expertise and personnel.

Even some of those scholars who are in general supporters of universities' entrepreneurial activities nevertheless have offered a cautionary note about the overhead expenses incurred by technology transfer programs (Geiger, 1992). Others point to the failure of institutions to deal in concert with costs and revenues. For example, Feller (1990) has warned against over investment in technology transfer programs. In part, his caution is linked to the failure of universities to accurately calculate costs as well as revenues.

There is almost no empirical work on the issue. One exception is a national survey of technology transfer offices in Britain. Packer and Webster(1996, p. 432), find that 60 percent of the respondents reported that "patenting activity was *not* self-financing." (emphasis in original)

Perhaps the best example of the type of work that is needed is Tolbert's (1985) use of resource dependency and institutional theory in studying various administrative offices. In the case of development offices, she finds that in the public sector where (at that time) development offices were not yet fully "institutionalized," the size of the office is correlated with the amount of funds raised as a share of institutional revenues. By contrast, in the private sector, where development offices are "institutionalized"—the normative environment is that every such institution should have one—there is no relationship between the size of the office and the amount of funds raised as a share of institutional revenues. In other words, growth in such offices is not simply a function of how much activity (or dependence) there is. There are other mechanisms at work.

The case of technology transfer professionals is instructive in this regard as well. It points to the fact that the state is more than regulator. The state is also a creator of opportunity structures. And rising/emergent professional groups take advantage of these structures to establish their positions and expand their offices. The growth of AUTM, and of technology transfer activity and professionals in universities, was really spawned by the Bayh-Dole Act of 1980, which enabled universities to own patents. Changes in state statutes and conflict of interest laws also enabled and encouraged the development of such offices on public university campuses (Slaughter and Rhoades, 1993). In this case, then, increased "administrative" costs were the result not of increased regulatory demands by the state, but of deregulation. The continued and further expansion of such managers (or as I have called them elsewhere, "managerial professionals"—Rhoades, 1996), is a function not of simple response to external pressure, but professional growth to fill a niche (opportunity structure) created by deregulation.

One could explain the growth of research administration in something of the same way. Indeed, an article published in the *SRA Journal* (the journal of the Society of Research Administrators) makes just this point:

> The growth in research administrators has definitely paralleled the in federal support of research since 1945. A recent check of the geographic distribution by state of members of the Society of Research Administrators and the National Council of University Research Administrators with the distribution of federal support to universities and selected nonprofit institutions by state showed a high positive correlation between the membership distribution of both organizations and the distribution of federal money. (Beasley, 1992, p.8)

The article goes on to suggest that the development and membership of three different associations of research administrators is also directly related to patterns of federal support for academic research. Of course, once these associations take root, they also take on a life of their own, driven in no small measure by concerns about professionalism, state of the art practice, and rationalizing and improving the quality of professional practice. As with faculty, the state's role in this branch of administration has not simply been one of regulator. It has also underwritten, spawned, and even encouraged the growth of research administration.

Conclusion

To summarize, then, I have suggested a variety of ways in which we should refine and rethink "administrative" costs. By way of refining, I stressed the need to disaggregate expenditure and personnel categories, and to move beyond (below and above) central administration. By way of rethinking, I emphasized the need to reconceptualize non-faculty personnel, not juxtaposing them against faculty as "administrative other," recognizing the restructuring of higher education's professional work force, addressing the changing nature of non-faculty professionals in such a way as to acknowledge their increasingly significant role as production workers, and focusing on their productivity. To clarify these points I drew on two empirical examples, unionized non-faculty personnel and technology transfer professionals.

Such refinement and rethinking, and the empirical examples I utilized, should lead scholars to undertake more sophisticated analyses of the location and causes of increased "administrative" costs. It should open scholars up to the various roles of "the state" in shaping the expansion of non-faculty professionals, directing attention not just to state regulations and reporting requirements, but to deregulation and to the sponsorship by various arms and agencies of the state of the growth of non-faculty professions. Perhaps most important of all, my reconceptualization is suggestive in terms of non-faculty occupations' own role in promoting expansion, in negotiating better terms of employment, in taking advantage of opportunity structures, and in "professionalizing," which generally means hiring additional personnel, proliferating "state-of-the-art" practices, and incurring additional costs.

In light of this pattern of non-faculty "professional" growth, it is all the more important to begin to think about accountability and assessment of productivity on the non-academic side of colleges and universities. There is much room for improvement here. For example, as late as 1989, very few AAU institutions even regularly reviewed non-academic units (Brown, 1989). In the early 1990s, there has been much talk and somewhat less action along the lines of Total Quality Management (Coate, 1993). Most of the focus of such efforts is on process efficiency. That is certainly related to productivity, but in a very incomplete way. For example, the emphasis is on process, not on investments and yields (or on the financial costs of the process).

There are a range of issues left to be addressed about non-faculty professional personnel, most of which in one way or another relate to "costs" and "productivity." For starters, we should begin to map the growth and professionalization of these occupations (see Rhoades, 1996). In addition, we should develop a fuller understanding of these managerial professionals' daily lives and everyday practices—"thick descriptions" of their work. If we are to make such employees more productive, we must first gain a better sense of exactly what it is that they do. Further, we should explore the social relations among these non-faculty professionals, and between them and faculty. The professional and political terrain of colleges and universities is far more complex than our current categories allow for. Such terrain has direct implications for how we can better organize our work and collective efforts.

In closing, my point is, here is a sector of the professional work force in higher education that is becoming increasingly significant in and central to the missions of colleges and universities. They are not just "ancillary" or "support" staff. Here is where the growth in higher education's work force lies. Here is where many of our students go. And we know very little about these non-faculty, managerial professionals.

REFERENCES

Andersen, K.E. (1991). Anatomizing bloat. *Academe* 77(6): 20-4.
Barley, S.R. (1996). Technicians in the workplace: ethnographic evidence for bringing work into organizational studies. *Administrative Science Quarterly* 41(3): 404-41.
Beasley, K.L. (1992: reprinted from 1971). A comparative analysis of the major research administrative organizations. *SRA Journal* 23(4): 7-15.
Bellas, M. (1994). Comparable worth in academia: the effects on faculty salaries of the sex composition and labor-market conditions of academic disciplines. *American Sociological Review* 59(6): 807-21.
Bergmann, B.R. (1991). Bloated administration, blighted campuses. *Academe* 77(6): 12-16.
Blau, P.M. (1973). *The Organization of Academic Work.* New York: Wiley.
Blau, P.M., and Scott, W.R. (1962). *Formal Organizations: A Comparative Approach.* Chicago: University of Chicago Press.
Bouchard, R., and Fortunato, R.T. (1990). Is your human resource department staffed adequately? *CUPA Journal* 41(2): 1-13.
Bouchard, R., Davidson, H.C., and Fortunato, R.T. (1992). Total organizational and personnel accounting study: a study of higher education human resources staffing in large institutions. *CUPA Journal* 43(1): 17-23.
Bowen, H.R. (1980). What determines the costs of higher education. In H.R. Bowen (ed.) *The Cost of Higher Education.* Washington, DC: Carnegie Foundation for the Advancement of Teaching.
Brinkman, P., and Leslie, L.L. (1986). Economies of scale in higher education: sixty years of research. *Review of Higher Education* 10(Fall): 1-28.
Brown, M.K. (1989). Developing and implementing a process for the review of non-academic units. *Research in Higher Education* 30: 89-112.
Caplow, T. (1957). Organizational size. *Administrative Science Quarterly* 1(March): 484-505.
Catell, N.O. (1985). The character of the university and the increasing threat of state government intrusion. In L.L. Leslie and H.L. Otto (eds.) *Financing and Budgeting Postsecondary Education in the 1980s.* Tucson, Arizona: Center for the Study of Higher Education.
Chavez, A. (1996). Strategies for survival in student affairs. Paper presented at annual meetings of Association for the Study of Higher Education, November, Memphis.
Coate, E. (1993). The introduction of TQM at Oregon State University. *Higher Education* 25: 303-20.
Dill, D.D. University-industry entrepreneurship: the organization and management of American technology transfer units. *Higher Education* 29(4): 369-84.
DiMaggio, P., and Powell, W. (1983). The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. *American Sociological Review* 48: 147-60.
Edelman, L.B., Abraham S.E., and Erlanger, H.S. (1992). Professional construction of law: the inflated threat of wrongful discharge. *Law and Society Review* 26: 47-84.
Fairweather, J. (1993). Academic values and faculty rewards. *Review of Higher Education* 17(1): 43-68.
Feller, I. (1990). Universities as engines of economic growth: they think they can. *Research Policy* 19(4): 335-48.

Friedman, R.S., and Friedman, R.C. (1984). Managing the organized research unit. *Educational Record* 65(Winter): 27-30.

Galambos, E.C. (1987). Higher education administrative costs and staffing. In Office of Educational Research and Improvement (ed.) *Higher Education Administrative Costs: Continuing the Study.* Washington, DC: U.S. Department of Education.

Geiger, R. (1992). The ambiguous link: private industry and university research. In W.E. Becker and D.R. Lewis (eds.) *The Economics of Higher Education.* Boston: Kluwer.

Geiger, R. (1990). Organized research units—their role in the development of university research. *The Journal of Higher Education* 61(1): 1-19.

Gilbert, S.W., and Green, K.C. (1995). *Information Technology: A Road to the Future?* Washington, DC: National Education Association.

Glaspar, R. (1995). Rising administrative costs in community colleges. Doctoral dissertation, University of Arizona.

Goode, W.J. (1957). Community within a community: the professions. *American Sociological Review* 22: 194-200.

Grassmuck, K. (1991). Throughout the 80s, colleges hired more non-teaching staff than other employees. *Chronicle of Higher Education,* August 14, p.22.

Grassmuck, K. (1990). Big increases in academic support staff prompt growing concerns on campuses. *Chronicle of Higher Education,* March 28, pp.1, 32-3.

Grunig, S.D. (1995). The impact of development office structure on fund raising efficiency for research and doctoral institutions. *The Journal of Higher Education* 66(6): 686-99.

Gumport, P.J., and Pusser, B. (1995). A case of bureaucratic accretion: context and consequences. *The Journal of Higher Education* 66,5: 493-520.

Gumport, P.J., and Pusser, B. (1994). Higher education funding shifts: privatization and the case of UC. Paper presented at the American Educational Research Association meetings, New Orleans, April.

Hackman, J.D. (1985). Power and centrality in the allocation of resources in colleges and universities. *Administrative Science Quarterly* 30(1): 61-77.

Halfond, J.A. (1991). How to control administrative cost. *Academe* 77(6): 17-19.

Hansen, W.L., and Guidugli, T. (1990). Comparing salary and employment gains for higher education administrators and faculty members. *The Journal of Higher Education* 61(2): 142-60.

Hawley, A.H., Boland, W., and Boland, M. (1965). Population size and administration in institutions of higher education. *American Sociological Review* 30(2): 252-5.

Hays, S.W. (1991). From adhocracy to order: organizational design for higher education research and service. *Research Management Review* 5(Spring): 1-17.

Heinz, J., and Laumann, E. (1982). *Chicago Lawyers: The Social Structure of the Bar.* New York and Chicago: Russell Sage Foundation and American Bar Association.

Hendershot, G.E., and James, T.F. (1972). Size and growth as determinants of administrative-productive ratios in organization. *American Sociological Review* 37(2): 149-53.

Hurd, R.W. (1995). *Directory of Staff Bargaining Agents in Institutions of Higher Education.* New York: The National Center for the Study of Collective Bargaining in Higher Education and the Professions, School of Public Affairs, Baruch College, City University of New York.

Indik, B.P. (1964). Organization size and supervision ratio. *Administrative Science Quarterly* 9(3): 301-12.

Klatzky, S.R. (1970). Relationship of organizational size to complexity and coordination. *Administrative Science Quarterly* 15(4): 428-38.

Leslie, L.L. and Rhoades, G. (1995). Rising administrative costs: seeking explanations. *The Journal of Higher Education* 66(2): 187-212.

Lunsford, T. (1970). Authority and ideology in the administered university. In C. Kruytbosch and S. Messinger (eds.) *The State of the University: Authority and Change.* Bev-

erly Hills: Sage Publications.
Mayfield, L.G., and Schultzman, E. (1987). Status report on the engineering research centers program. *Research Management* 30(1): 35-41.
Meyer, J.W., and Rowan, B. (1977). Institutionalized organizations: formal structure as myth and ceremony. *American Journal of Sociology* 83: 340-63.
Meyer, M., Stevenson, W., and Webster, S. (1985). *Limits to Bureaucratic Growth.* Berlin: Walter de Gruyther.
Montgomery, D., and Lewis, G. (1995). Administrative staff: salaries and issues. *National Education Association 1995 Almanac of Higher Education.* Washington, DC: National Education Association.
Montgomery, D., and Lewis, G. (1994). Administrative staff: salaries and issues. *National Education Association 1994 Almanac of Higher Education.* Washington, DC: National Education Association.
Mortimer, K.P., and McConnell, T.R. (1978). *Sharing Authority Effectively.* San Francisco: Jossey-Bass.
Packer, K., and Webster, A. (1996). Patenting culture in science: reinventing the scientific wheel of credibility. *Science, Technology, and Human Values* 21(4): 427-53.
Parkinson, C.N. (1957). *Parkinson's Law and Other Studies in Administration.* Boston: Houghton Mifflin.
Parson, T. (1954). *Essays in Sociological Theory.* New York: Free Press.
Pondy, L.R. (1969). Effects of size, complexity, and ownership on administrative intensity. *Administrative Science Quarterly* 14: 47-60.
Powell, W., and DiMaggio, P. (eds.) (1991). *The New Institutionalism in Organizational Analysis.* Chicago: University of Chicago Press
Rhoades, G. (forthcoming). *Managed Professionals: Unionized Faculty and Restructuring Academic Labor.* Albany: State University of New York Press.
Rhoades, G. (1996). Reorganizing the work force for flexibility: part-time professional labor. *The Journal of Higher Education* 67(6): 626-59.
Rhoades, G. (1996). Managerial professionals: the rise and restratification of professionals in higher education. Paper presented at annual meetings of Association for the Study of Higher Education, November, Memphis.
Rhoades, G. (1995a). Rising, stratified administrative costs: student services' place. In D.B. Woodard, Jr. (ed.), *Budgeting as a Tool for Policy in Student Affairs.* New Directions for Student Services, Number 70. San Francisco: Jossey-Bass.
Rhoades, G. (1995b). Rising administrative costs in instructional units. *Thought and Action* 11(1): 7-24.
Rhoades, G. (1993). Retrenchment clauses in faculty union contracts: faculty rights and administrative discretion. *The Journal of Higher Education* 64(3): 312-47.
Rhoades, G., and Slaughter, S. (1994). From balance wheel to cogwheel: the NSF from 1976-1992. Paper presented at the Association for the Study of Higher Education meetings, Tucson, November.
Rhoades, G., and Slaughter, S. (1991). Professors, administrators, and patents: the negotiation of technology transfer. *Sociology of Education* 64(2): 65-77.
Scott, W. R. (1966). Professionals in bureaucracies—areas of conflict. In H.M. Vollmer and D.L. Mills (eds.) *Professionalization.* Englewood Cliffs, NJ: Prentice-Hall.
Selznick, P. (1969). *Law, Society, and Industrial Justice.* New York: Russell Sage Foundation.
Simpson, W.B. (1991). *Cost Containment for Higher Education: Strategies for Public Policy and Institutional Administration.* New York: Praeger.
Skrentny, J.D. (1996). *The Ironies of Affirmative Action: Constructing Justice in a Turbulent America.* Chicago: University of Chicago Press.
Slaughter, S. (1993). Retrenchment in the 1980s: the politics of prestige and gender. *The*

Journal of Higher Education 64(3): 250-82.

Slaughter, S., and Rhoades, G. (1996). The emergence of a competitiveness research and development policy coalition and the commercialization of academic science and technology. *Science, Technology, and Human Values* 21(3): 303-39.

Slaughter, S., and Rhoades, G. (1993). Changes in intellectual property statutes and policies at a public university: revising the terms of professional labor. *Higher Education* 26: 287-312.

Smart, J.C. (1991). Gender equity in academic rank and salary. *Review of Higher Education* 14(4): 511-26.

Stahler, G.J., and Tash, W.R. (1994). Centers and institutes in the research university: issues, problems, and prospects. *The Journal of Higher Education* 65(5): 540-54.

Stanley, E.C., and Adams, J.W. (1994). Analyzing administrative costs and structures. *Research in Higher Education* 35(1): 125-40

Summerville, R.M., and Ridley, D.R. (1992). Toward establishing salary benchmarks for college and university administrators. *CUPA Journal* 43(4):17-23.

Sutton, J.R., and Dobbin, F. (1996). Responses to legal uncertainty in U.S. firms, 1955 to 1985. *American Sociological Review* 61(5): 794-811.

Teich, A. (1982). Research centers and non-faculty researchers: a new academic role. In D.I. Phillips and B.S. Shen (eds.), *Research in the Age of the Steady-State University.* AAAS Selected Symposium Series, Boulder, Colorado: Westport Press.

Tolbert, P. (1985). Institutional environments and resource dependence: administrative structure in institutions of higher education. *Administrative Science Quarterly* 30: 1-13.

Townsend, B.K., and Mason. S.O. (1990). Career paths of graduates of higher education doctoral programs. *Review of Higher Education* 14(1): 63-82.

Zemsky, R., and Massy, W.F. (1990). Cost containment. *Change* 22 (November-December): 16-22.

Zumeta, W., and Looney, J. (1994). State policy and budget development. *National Education Association 1994 Almanac of Higher Education.* Washington, DC: National Education Association.

State Efforts to Keep Public Colleges Affordable in the Face of Fiscal Stress[1]

Michael Mumper
Ohio University

After more than a decade of relative stability, the price of public colleges and universities began to rise rapidly in the early 1980s. While these price increases slowed somewhat in the last half of the 1980s, rapid tuition inflation returned in the 1990s. Today, public higher education in most states is more expensive than it has ever been. As college prices rise, so too does public concern over them. Few issues now concern American families as much as the spiraling price of college (Gallup, 1991). This combination of rising prices and public concern over them has, in turn, generated substantial pressure on policy makers to take action to address the problem.

Most of the public debate over declining college affordability has focused on the astronomical costs of the most expensive private colleges. Certainly, few American families can afford the full $30,000 per year now charged at many of these institutions without depleting their retirement savings, remortgaging their homes, or going far into debt. However, much less attention has been directed at the growing problem of public college affordability. In 1980, states like California, New York, North Carolina, and Texas provided public higher education to state residents for virtually no charge. Even the public colleges in high tuition states like Pennsylvania, Ohio, and New Hampshire were bargains by today's standards. But after nearly two decades of tuition inflation, prices have risen to the point that average public college tuition in several states is more than $6,500 a year. When living expenses and books are included, the price tag for four years at a public university can easily exceed $40,000 per student.

Beginning in 1965, with the enactment of the Higher Education Act, the federal and state governments joined in an implicit partnership to make higher education affordable to all Americans. The states would keep tuition at public colleges low and the federal government would provide a means-tested system of

[1] I would like to acknowledge the support of the Ohio University Research Committee in completing the research for this chapter. The project was also aided substantially by the careful and conscientious research assistance of Lisa Eiserman.

J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*

grants and loans to insure that all students could meet those costs. In recent years, however, that partnership has come unraveled. Faced with public pressure to reduce taxes and balance the budget, national policy makers have frozen the size of the grant programs and forced more and more students to rely on loans to finance their higher education (Gladieux and Hauptman, 1995). As the federal government has backed away from its traditional responsibility for insuring college access for the lower income and disadvantaged, that role has been largely given over to the states and to families. Today, it is clear that if the nation's public colleges are to remain affordable to all Americans—even the lowest income families—it will be the result of actions taken by state level policy makers.

The states, however, have followed quite different paths in their efforts to insure equal college opportunity. This chapter will examine how state governments have addressed this complex problem in the 1990s[2]. Specifically, it will consider:

- 1) Why have public college prices increased so rapidly? And what accounts for the variation among states both in tuition levels and in the levels of tuition inflation?
- 2) How have states responded to these pressures? What actions have they taken to try to insure that their public colleges will be affordable to lower income and disadvantaged students?
- 3) Which state actions have proven the most effective? As states struggle to do more with fewer resources, which approaches seem to provide the greatest hope for improving college affordability?

The Public College Affordability Problem[3]

While public college prices have increased everywhere in the past decade, those increases have varied tremendously across the states and across time. Table One shows the average tuition at four-year public colleges for selected years between 1984-85 and 1994-95. It shows that all but five states have experienced tuition inflation of greater than 100 percent during the decade. The states with the fastest rates of increase (California, Texas, Massachusetts, and Connecticut) were all hit hard by the recession of the early 1990s and were among the last to recover. Moreover, because California and Texas had maintained such low tuition at the beginning of the period, their price increases of 557 and 318 percent, respectively, still did not move them to among the most expensive states. Indeed, in spite of this dramatic price increase, Texas remains among the states with the least expensive public colleges.

---

[2]The focus of this chapter will be limited to the affordability of undergraduate education at public institutions. The issues of access and affordability to graduate professional and private institutions of higher education are certainly important. But they are substantially different than those discussed here.

[3]Data in this section are drawn from the *Digest of Education Statistics 1995* (Washington, DC: Government Printing Office).

TABLE 1. Four-year Public College Tuition

State	1984-1985	1989-1990	1994-1995	5 yr. percent change	10 yr. percent change
California	411	1,123	2,703	173	558
Texas	384	959	1,608	150	319
Massachusetts	1,130	2,052	4,131	82	266
Connecticut	1,044	2,017	3,746	93	259
Wyoming	567	1,003	1,908	77	237
Washington	849	1,710	2,686	101	216
New Jersey	1,225	2,511	3,773	105	208
North Carolina	494	1,015	1,503	106	204
Illinois	1,060	2,370	3,197	124	202
Oregon	1,024	1,738	3,063	70	199
Missouri	944	1,532	2,787	62	195
Maryland	1,175	2,120	3,318	80	182
Rhode Island	1,322	2,281	3,718	73	181
Virginia	1,345	2,532	3,769	88	180
Arizona	680	1,362	1,894	100	179
South Carolina	1,085	2,162	3,021	99	178
Michigan	1,368	1,484	3,729	9	173
Delaware	1,405	2,768	3,817	97	172
Louisiana	816	1,768	2,214	117	171
Mississippi	903	1,858	2,448	106	171
Alaska	768	1,280	2,045	67	166
Oklahoma	631	1,309	1,675	107	166
New Mexico	717	1,326	1,836	85	156
Montana	833	1,535	2,110	84	153
Hawaii	596	1,293	1,508	117	153
Pennsylvania	1,870	3,210	4,512	72	141
Vermont	2,427	3,641	5,752	50	137
Florida	759	na	1,786	na	135
Kentucky	879	1,316	2,056	50	134
Tennessee	812	1,406	1,897	73	134
Arkansas	837	1,376	1,955	64	134
Maine	1,425	1,980	3,319	39	133
Alabama	911	1,522	2,106	67	131

TABLE 1. Four-year Public College Tuition *(continued)*

State	1984-1985	1989-1990	1994-1995	5 yr. percent change	10 yr. percent change
Iowa	1,082	1,823	2,462	69	128
Wisconsin	1,087	1,861	2,470	71	127
North Dakota	992	1,604	2,245	62	126
West Virginia	886	1,591	1,963	80	122
New Hampshire	1,814	2,196	4,003	21	121
New York	1,355	1,460	2,957	8	118
Ohio	1,577	2,432	3,405	54	116
Kansas	954	1,467	2,019	54	112
DC	496	664	1,046	34	111
Indiana	1,371	1,975	2,864	44	109
South Dakota	1,243	1,718	2,557	38	106
Colorado	1,162	1,830	2,377	57	105
Utah	974	1,429	1,960	47	101
Minnesota	1,500	2,063	2,919	38	95
Nebraska	1,059	1,519	2,058	43	94
Georgia	1,084	1,631	1,965	51	81
Idaho	889	1,119	1,583	25	78
Nevada	915	1,100	1,601	20	75
U.S. Total	971	1,781	2,689	84	177

Source: U.S. Department of Education (1995)

At the other end of the spectrum, Nevada, Idaho, and Georgia were the states with the lowest rates of tuition inflation. In the case of Georgia and Nevada, this is the result of both a healthier than average state economy and a growing population base.

Table One illustrates that in spite of some high profile news accounts of a leveling off of tuition inflation, the price of public colleges increased more rapidly in the last half of the decade than it had in the first. New York, Michigan, Maine, and Minnesota, in particular, were able to maintain their low levels of tuition inflation between 1985 and 1990. But each experienced dramatically greater growth in the following years. New York, in particular, was able to maintain tuition increases at its four-year colleges to less than 10 percent in the last half of the 1980s (Mumper, 1993). But prices there more than doubled between 1990 and 1995.

Table Two shows average tuition at two-year public colleges by state between 1990 and 1995. The pattern here is the same. Every state has experienced tuition increases in the last five years, although the rate of inflation in this sector remained substantially lower than it was among four year colleges. Again, those states with the highest rates of tuition inflation were either those, like California, Maine, and Massachusetts, which were hit especially hard by the recession and/or states like North Carolina and California where 1990 tuition levels were far below those charged in other states.

TABLE 2. Two-year Public College Tuition

State	1989-1990	1994-1995	5 yr. percent change
California	112	365	226
North Carolina	288	582	102
Maine	1,134	2,137	88
Massachusetts	1,332	2,441	83
Oregon	753	1,324	76
Virginia	813	1,384	70
Alabama	662	1,123	70
Rhode Island	1,004	1,686	68
Connecticut	915	1,520	66
Washington	802	1,314	64
West Virginia	803	1,312	63
Colorado	792	1,279	62
Nevada	522	842	61
Maryland	1,172	1,848	58
Kentucky	693	1,080	56
New Jersey	1,130	1,755	55
Florida	729	1,112	53
New York	1,412	2,152	52
Montana	877	1,329	52
Texas	455	680	50
Missouri	815	1,203	48
Kansas	711	1,044	47
Wyoming	613	893	46
New Hampshire	1,608	2,316	44
Delaware	882	1,266	44
Wisconsin	1,160	1,649	42
Arizona	519	734	41
Iowa	1,225	1,699	39
Arkansas	644	888	38
Mississippi	680	935	38

TABLE 2. Two-year Public College Tuition *(continued)*

State	1989-1990	1994-1995	5 yr. percent change
Oklahoma	840	1,153	37
Illinois	871	1,194	37
Michigan	1,047	1,432	37
New Mexico	496	678	37
Vermont	2,120	2,877	36
Indiana	1,374	1,854	35
Ohio	1,636	2,164	32
North Dakota	1,286	1,689	31
South Carolina	807	1,048	30
Minnesota	1,499	1,928	29
Idaho	779	990	27
Pennsylvania	1,419	1,751	23
Louisiana	837	1,027	23
Hawaii	410	500	22
Tennessee	803	975	21
Nebraska	919	1,097	19
Georgia	852	1,015	19
Utah	1,136	1,340	18
Alaska	na	1,320	na
South Dakota	na	3,430	na
U.S. Total	758	1,194	58

Source: U.S. Department of Education (1995)

While, in general, the trends in four-year and two-year college tuition inflation are moving in the same direction, there are a few notable exceptions. Both Illinois and South Carolina have held tuition increases at their two-year colleges substantially below those experienced at four-year colleges. By holding down tuition inflation at two-year colleges, these states can insure that lower income families have access to more affordable public higher education and still allow their research universities to generate the revenues which result from higher tuition. A similar approach has been used in California which has always kept tuition at two-year colleges very low. This represents an implicit strategy to maintain college affordability by encouraging students to attend institutions which provide an education at the lowest price.

A great deal can be learned from examining trends in aggregate tuition rates in the states. But these figures may also mask important differences in public college affordability. Tuition at public colleges within a state may vary widely among institutions. This may even occur within a single sector. As such, the average tuition may not accurately represent the amount most students must pay. Some students will pay more than the average. Others will pay much less.

Perhaps more important, low tuition is not the same as affordability. By providing need-based financial aid, states can substantially reduce the real price of college for lower income students. If the rising tuition in a state is paid only by more wealthy students, those increases may have little or no impact on the college prices paid by the most disadvantaged. Complicating matters further, institutions themselves offer financial aid that may also mitigate the impact of tuition increases. As such, rising tuition cannot be equated with declining affordability in a state. And even stable tuition is no guarantee that public college access for lower income students is not declining.

The States and Public Higher Education

The fact that public college tuition has increased rapidly in recent years is undeniable. The causes of these increases, however, remain the subject of much dispute. Some see rising prices as the product of insufficient support from the federal and state governments (Hauptman, 1990, p. 59-64). Some see them as driven by wasteful and unnecessary expenditures by campus leaders (Sowell, 1992). Some see them as the product of bloated bureaucracies and red tape (Bergmann, 1991). Still others see rising college prices as a direct result of the rising costs of purchasing and maintaining new technology, providing health care for employees, and covering the costs of complying with federal regulations (Francis, 1990). As these costs of offering a quality education rise, institutions raise their prices in order to balance their budgets.

Putting aside these disagreements over causes, there are three things that all sides agree have played a role in causing public college prices to rise. First, due to the fiscal pressures they have experienced in the past few years, states have been unable to significantly increase their appropriations to higher education. Steven Gold (1995), one of the most careful observers of state spending patterns, found that in the early 1990s "higher education took a worse beating than any other spending category" (p. 25) in state budgets. In several states, the level of support has actually declined. In Virginia, for example, state support for higher education fell by $500 million or 27 percent between 1990 and 1995 (Hsu, 1995). Second, public colleges have increased their spending on such things as student services, computing facilities, and faculty and administrative compensation. Finally, as more students are going to college, and more of them begin with academic deficiencies, colleges are spending more time and money providing remedial courses to prepare students to do introductory college work. This combination has forced college leaders to provide more students with a wider array of services, to pay higher prices to provide those services, and to do so with stable or reduced levels of state support. The result has been tuition inflation.

But there is something more than just these forces at work. There is also evidence of changing state priorities. Several states seem to have decided to shift the responsibility of paying for higher education from taxpayers to students and their families. Sandra Ruppert puts it this way:

> To some extent, high tuition values reflect the state's position that the individual is the primary beneficiary of his or her education and so students and their families should bear more of the cost. Conversely, high state appropriations suggest a state's position that higher education provides social and economic benefits for states and localities so government should bear more of the cost (Ruppert, 1996, p. 35).

State Government's and Public College Prices

When viewed in aggregate, state spending on higher education dwarfs the combined efforts of federal and local governments. States provide most of these dollars directly to public colleges and universities as instructional subsidies. Colleges, in turn, use these subsidies to keep their tuition considerably below the full cost of providing a higher education. Another large portion of those state funds is allocated for need-based grants which are awarded directly to lower family income students. These grants are given with the express purpose of providing students with the resources necessary to attend the college of their choice. As such, the level of state support is a central factor in the affordability of public higher education in a state.

While state governments are major players in the financing of public higher education, the relationship between state governments and public colleges is complex and multifaceted. While states provide more than half of the revenues currently used to operate public institutions of higher education, these colleges retain substantial autonomy in their administrative and financial operations. In most states, it is the colleges themselves, acting through boards of trustees, who determine the tuition they will charge.[4]

In only a few states, like New York, Texas, and Washington, does the legislature actually set the tuition level charged at public colleges. But even in these states, it is usually set in careful consultation with the state's board and college presidents (Lenth, 1993).

There is a clear relationship between levels of state support and the tuition charges at a state's public colleges. Public colleges receive their funding from a combination of state support, tuition, private contributions, and sales and services. When state support declines, colleges plug the revenue gap by increasing tuition charges. That is exactly what has happened since 1990. As Thomas Mortenson (1994) describes it:

> As states' governors and legislators have chosen to shift state appropriations away from higher education into more "important" budget priorities like corrections and Medicaid, public institutions have raised tuition charges to students to offset the loss of state appropriations (p. 7).

---

[4]For a discussion of the different approaches, see Lenth, 1993.

A 1996 study by the National Education Association (NEA), which reports the findings of a survey of the chairs of education committees in state legislatures, makes a similar point. It found that a part of the reason higher education has been a low budgetary priority is that the burden of these cuts can be easily shifted. Many legislators are aware that a share of the cuts to the higher education appropriation can be shifted to others primarily in the form of tuition and fee increases. Indeed, a strong majority of committee chairs (68 percent) agreed that a significant factor on determining how much money the legislature will appropriate for higher education is the ability of colleges and universities to raise their own money through tuition, research grants, and gifts (Ruppert, 1996, p. 9).

While the link between state appropriations and tuition at public colleges is a real one, it is far from lockstep. The determination of the tuition level charged each year at public colleges is the product of countervailing pressures. Long traditions of institutional autonomy from direct government control mandate that campuses retain at least some control over expenditure and pricing decisions. But growing public pressures for accountability in the use of public funds demand that state governments act to control or limit price increases (Berdahl and McConnell, 1994). Whether or not those public pressures are translated into explicit governmental action, campus decision makers are certainly aware that rapidly rising tuition makes them the target of increased scrutiny by governors and state legislatures (Lenth, 1993, p. 7-15). As such, public colleges must establish their own price levels within the context of limited state appropriations and the need to be responsive to broad public and political pressures.

The Fiscal Condition of the States

The 1990s have been a difficult time for state governments. Gold (1995) recently termed the period "the fiscal crisis of the states." The most important cause of the fiscal problems faced by the states was the recession of the early part of the decade. This created a squeeze in which state tax revenues were declining as the demand for many public services was increasing. Although the national recession ended in early 1991, many states experienced its effects for a much longer time. A few appear to have not yet recovered.

While important, the recession was not the only cause of the fiscal problems faced by the states in the 1990s. The explosive growth of Medicaid and the increased impact of federal mandates required all states to spend more in these areas without generating any additional revenue to cover those costs (Miller, 1993). Many states experienced significant increases in public school enrollment or faced court orders to reduce disparities in spending among school districts (Zumeta and Looney, 1993, p. 8-9). Similarly, mandatory sentencing requirements and tougher practices in the criminal justice system forced states to greatly increase their spending on corrections and prison construction (Gold, 1995, p. 27).

The ability of a state to support its public colleges is a function of both the state's capacity to raise revenue and the willingness of the legislature to appropriate funds. Edward Hines (1996) describes it this way:

> The amount of revenue available depends on the capacity of the tax system and the overall health and level of activity in the economic system. State appropriations take place in a system characterized by tradeoffs and choices. At a time when states' finances are under strain by increasing demands on state services, the willingness of lawmakers to appropriate to one area may mean lawmakers will be unable to support another area of need (p. 6).

In combination, these factors have forced states to reallocate the shares within their annual budgets. As shown in Table Three, since 1990 there were increases in all of the major state expenditure categories except higher education. This is no accident. State policy makers were trapped between pressures to increase spending on K-12 education, prisons, medical care, and welfare on the one hand and pressures to hold down taxes and legal requirements to balance their state budgets on the other. Given these cross-pressures, many policy makers felt as though higher education was the only place they could reduce spending without producing a short-term disaster (Ruppert, 1996, p. 9). As such, even when not accounting for inflation, state spending on higher education experienced an annual decline of 0.6 percent per year between 1990 and 1995.

TABLE 3. Mean Annual Change in Major Expenditure Categories from State General Funds: 1990-1995

Major State Expenditures	Mean Annual Change
Medicaid	10.0%
Prisons	8.5%
K-12 Education	3.7%
AFDC	1.6%
Higher Education	-0.6%

Source: Mortenson, 1994.

These trends seem likely to remain evident at least through the end of the century. Even in those states which are experiencing rapid economic growth, the need for additional spending on health care, prisons, and elementary and secondary education will continue to attract the bulk of the new resources. In those states where the economy is weak, higher education is likely to experience disproportionate reductions (McGuinness, 1994, p. 159).

The Specter of Enrollment Increases

Complicating matters further, many states are facing the forecast of an increasing demand for higher education in the next decade. These states, many of which are already operating their systems at or near full capacity, must find ways to accommodate these new students without the funds necessary to open new campuses or dramatically expand the physical capacities of their existing campuses (Ruppert, 1996, p. 27-28).

This new student demand is coming from two places. First, the number of nontraditional students returning to school has been growing for more than two decades. This group now significantly outnumbers the traditional 18-21 year old now enrolled in American higher education. The vast majority of these nontraditional students are enrolled in public colleges and universities. During the last decade, public colleges were able to absorb this new demand because the number of high school graduates was declining. Thus, even as the participation rates of high school graduates was growing, their absolute number in colleges was increasing more slowly.

But over the next decade, the number of high school graduates in many states will increase rapidly. As shown in Table Four, ten states will have an increase in the number of graduates of more than 30 percent. If these new graduates plan to attend college at the same rates as current graduates (and forecasts are that they will), and the number of nontraditional students continues to grow (and forecasts are that they will), the challenge will be enormous. How will states continue to provide affordable higher education to all their residents in the face of rising costs, scarce resources, and rapidly expanding demand? In many states the answer will be higher tuition at the public colleges.

TABLE 4. Projected Changes in the Number of High School Graduates Between 1995-1996 and 2005-2006

Nevada	71%
Florida	51%
Arizona	46%
California	43%
New Hampshire	36%
Alaska	33%
Colorado	33%
Delaware	32%
Washington	32%
Maryland	30%

Source: Ruppert (1996). Appendix C: State Date Table, pp. 57-58.

The Policy Responses of the States

States have not been passive in the face of the growing strain on public college affordability. Voters, taxpayers, parents, and students all demand that states take action to keep public colleges affordable. As a result, policy makers in most states have been forced to respond, in one way or another, to those pressures. Their responses, however, have varied widely. Few states have developed a comprehensive strategy or policy to control college prices (Hearn, Griswold, and

Marine, 1996). Most states have undertaken piecemeal, ad-hoc, incremental changes on several fronts simultaneously. In some states the response came directly from the state legislature or the governor's office. In others, it came from the Board of Regents or the state's higher education coordinating board. In others, it came as informal pressure on campus leaders from state policy makers to induce them to change their behavior or policies.

In general, the state responses can be divided into four categories. States have sought to (1) alter the expenditure patterns of public colleges, (2) increase the revenues available to public college leaders to offset their rising costs, (3) redesign the delivery of higher education in order to make it more efficient and thus reduce its cost, and/or (4) increase the resources available to students and their families to pay for those costs. The next section of this chapter looks more closely at these responses. It is important to remember, however, that these are not mutually exclusive categories. States can choose to respond in a variety of ways at once. Many states have done a little of each. Other states have concentrated their efforts in one or two of them.

RESPONSE #1: Controlling Institutional Expenditures

As state policy makers struggle to address more and often more serious problems with a limited pot of funds, they often begin by attempting to insure that existing funds are spent in the most productive way. This pressure to do more with less has caused legislatures and governors to demand increased accountability from their public colleges. In most cases, states want to insure that colleges are spending their state funds, and also the funds they generate from tuition, in ways that they feel are serving the state's interests (Mathesian, 1995).

These efforts have renewed old tensions between states and campuses over institutional autonomy. To some, they may have fundamentally altered the relationship between capital and campus. William Shkirti, Vice President for Finance at Ohio State University, puts it this way: "the assumption used to be that higher education was a good investment...now we're being asked to prove that we're not inefficient" (Mathesian, 1995, p. 21).

Mandated Price Controls

The ways public college prices are set vary widely from state to state. But in almost every state, the legislature and the governor can, if they choose, exert a powerful influence on the process. Indeed, legislatures in almost every state have the power to freeze tuition increases either directly by statute or indirectly through appropriations negotiations. By taking actions to limit the tuition charged by public colleges, states are indirectly limiting the revenue available to those colleges to provide educational programs and services. Indeed, by limiting tuition increases, states are implicitly ordering campus leaders to change their spending priorities.

Legislative action to limit public college price increases is especially easy in states where public college tuition is actually set by the state legislature. By keeping tuition very low, many state legislators feel that they maintain the widest possible access to their public colleges. But low tuition generates little revenue to cover educational costs and state legislatures often provide colleges with little additional revenue to cover their rising costs. As a consequence, public college leaders in these states have complained bitterly that they are being squeezed in such a way that they have little choice but to compromise program quality.

In Texas the low tuition policy is more complex than it might appear. To insure that colleges have adequate revenues, and also maintain their policy of very low tuition, the state legislature and public colleges have reached an uneasy compromise. Colleges have been allowed to establish a number of student fees and charges which cover the cost of a specific activity or service. These fees are then used to supplement the campus budget and serve to replace the funds which are not appropriated by the state. For many, perhaps most, Texas students, these fees are now larger each term than their total tuition bill (Texas House Research Organization, 1995).

In states where tuition is set by a state board or by individual campuses, the process of state control is slightly more difficult. In Virginia and Ohio, for example, the legislature has established annual tuition caps for all public institutions. These caps, which were set at the level of consumer inflation in Virginia and slightly higher in Ohio, establish the maximum percentage of tuition increase that the state will accept each year. In both states frustrated legislators saw tuition caps as a way to show campus leaders they were serious about controlling costs. In the view of many legislators, this was the only way to force unresponsive campuses to limit their wasteful spending and set reasonable priorities. When coupled with reductions in state appropriations, legislators felt that such caps were the only way to insure that necessary belt-tightening took place on campus and the budget cuts were not simply passed on to students in the form of higher tuition.

Few people see state mandated tuition freezes as a long term solution to the college affordability problem. They are simply a way for state policy makers to force campus leaders into adjusting their spending priorities. Often such actions are the first step in a renegotiation of the relationship between campus and state leaders. Ohio State Representative Robert Hagen puts it this way, "those in the ivory tower have to come down, get a little muddied and explain the whole process of what it is they do" (Mathesian, 1995, p. 22).

Performance Funding

A more positive way for states to alter the spending patterns of public colleges is performance funding. This approach ties state funding levels to measures of institutional performance. In theory, high performing institutions will be rewarded with funding increases while low performing institutions will be punished with funding reductions (Nedwek, 1996).

During the 1980s, a few states developed "incentive funding" programs that added dollars to the base budgets of institutions that implemented an approved plan that fit state priorities. Tennessee was the first state to adopt such a program, followed by Florida, New Jersey, and Ohio (Zumeta, 1995, p. 83). These incentive plans link increased funding with improving undergraduate teaching, increasing classroom use of technology and improved job training at community and technical colleges among others. In Colorado, the legislature passed an incentive funding plan which rewards colleges and universities that improve in five statewide priority areas: productivity, growth in enrollment, expanded job training, strengthened ties to public schools, and increased financial aid (Colorado Commission on Higher Education, 1994).

In 1993, the Texas legislature was poised to enact a performance-based budgeting system to distribute 5 percent of the state's higher education funds. But when the plan was revised to increase the share to 10 percent, the consensus behind it evaporated. It died in committee and has never been revived (Mathesian, 1995, p. 24).

In South Carolina the legislature has taken a much more comprehensive approach to performance funding. The state's higher education commission developed a set of performance indicators for each state college. By 1999, the plan would base all of the funds it gives to each public college on how well they perform according to those indicators (Schmidt, 1996).

These efforts to improve the accountability of public higher education through performance funding are not directly related to the problem of rising tuition. They do, however, illustrate significant state efforts to alter the funding priorities of public colleges. In general, performance funding seeks to reward institutions that focus on undergraduate education, effective teaching, and job preparation (Zumeta, 1995). In doing so, states are implicitly discouraging more spending on research, public service, and graduate education.

The survey of education committee chairs conducted by the NEA shows substantial support for performance measures in funding higher education. Nearly half of the respondents think that their legislatures are likely in the next few years to "link funding to campus efforts to increase enrollment, graduation rates, or other measures of student or institutional performance" (Ruppert, 1996, p. 37).

Improved Efficiency and Productivity

Another way for states to alter the expenditure patterns of campus leaders is to mandate improved efficiency and productivity. While the goals of these mandates are similar to performance funding, they are much more intrusive. Rather than simply measuring performance on predetermined criteria, governors and legislators can simply require that colleges change their spending priorities. These state efforts can range from mandated increases in faculty workloads to adjustments in the types and availability of courses offered. In some states, legislatures have overhauled campus purchasing procedures and/or mandated reductions in specific spending categories. In other states, these efforts were more broadly designed to bring campus spending in line with available revenues and reduce the need for tuition increases.

Virginia has been among the most aggressive states in their efforts to improve efficiency and productivity. In 1994, the Governor appointed a Commission on Government Reform to search for ways to improve efficiency across the entire state government. This Commission ominously called itself the Blue Ribbon Strike Force. As part of their work, and with the support of the General Assembly, the Commission required all public colleges and universities in Virginia to submit extensive restructuring plans:

> to effect long term changes in the deployment of faculty, to insure the effectiveness of academic offerings, to minimize administrative and instructional costs, to prepare for the demands of enrollment increases, and to address funding priorities as approved by the General Assembly (Virginia Council on Higher Education, 1994, p. 1).

These plans were to follow strict criteria developed by the state's Council on Higher Education.

As part of this process, each public institution was required to conduct a comprehensive review with an eye toward reducing programs, eliminating expenses, and saving money. While the language of the plan stresses quality enhancement, participants in the process viewed the focus of the plan as on improving efficiency and reducing the size and scope of the higher education system (Lively, 1995). The entire process was met with stiff opposition from college and business leaders in the state (Hsu, 1995).

One institution where this process had a significant impact was James Madison University. After conducting the state mandated review, the University's president recommended the elimination of the physics major. His reasoning was that as institutions more clearly define their missions, not every college needed to offer every program. Indeed, the purpose of the review had been to eliminate weak and unnecessary programs. While in the end, the physics major was restored, the battle illustrated the determination of the state to take dramatic action to reduce spending on higher education (Magner, 1995).

A similar study was conducted in Ohio between 1990 and 1993. Under orders from the state legislature, the Ohio Board of Regents examined ways to reduce costs and prepare for the future needs of Ohio students and taxpayers. The report focused on the need to reduce duplication of high cost programs, increase cooperation between sectors of the higher education system, and increase faculty productivity (Cage, 1995). Perhaps its most controversial recommendation was to require all public colleges to increase faculty teaching loads by 10 percent. However, measurement of what constituted teaching, and how the requirement would be enforced, was left to campus officials (Tucker and Voelker, 1995).

In Colorado, the Commission focused its attention on the need to accommodate projected enrollment increases. But unlike in Ohio, the Commission's proposals were designed around positive incentives rather than threats and mandates.

Their recommendations included steps to encourage students to choose community college over research universities, and the elimination of subsidies for remedial courses and courses which were avocational rather than academic or vocational in nature (Colorado Commission on Higher Education, 1994).

These efforts by legislators and governors to alter the spending patterns of public colleges are symptoms of a broader tension. Campus and state leaders often have starkly different views of the appropriate priorities and practices. In particular, legislators began to question how faculty members spend their time. In 1993-94, 24 states conducted studies of faculty workload or productivity (Chronicle, 1994). As Daniel Layzell (1996) describes it:

> The logic in the minds of state legislators is clear: the more time faculty spend in the classroom, the more undergraduate students who can be accommodated at (lower cost), and the higher quality education received by those students (p. 164).

This concern over faculty teaching loads represents a fundamental tension between state lawmakers and campus leaders. The NEA survey of education committee chairs found that 86 percent felt strongly that college and university faculty should focus more on undergraduate education. The same survey found that 67 percent of legislators think that college faculty should teach more courses. Further, more than one in four education committee chairs think that their legislature will take action to mandate higher teaching loads in the next three to five years. As long as the gulf between the views of campus administrators and state leaders remains so wide, states are likely to continue to seek ways to alter campus spending patterns to more closely fit state priorities.

RESPONSE #2: Increasing Institutional Revenues

Rather than focusing on campus expenditures, some states have responded to increasing college prices by making adjustments in the revenue streams available to campus leaders. In its most straightforward manifestation, this means increasing the state appropriations to public colleges. By providing colleges with additional revenues, states can reduce the fiscal squeeze facing campus leaders and lessen the need for tuition increases. Such a response makes sense if the root cause of tuition inflation is that colleges have insufficient resources to provide a quality education at a stable price.

From a state's perspective, increasing institutional revenues is a very costly solution to the problem of tuition inflation. Even in states where policy makers see insufficient state support as the cause of rising college prices, limited state resources and shifting state priorities make increasing instructional subsidies unrealistic. There is simply not enough money to increase spending on higher education and still meet other state demands.

Generally, states build each year's budget by making incremental adjustments to last years budget. As such, the level of budgetary increase or reduction experienced by higher education often is a function of nothing more than how much money the state legislature has to allocate. Sandra Ruppert (1966) quotes one legislator who put the situation clearly "the most significant factor in whether we appropriate more money is if we have more money to appropriate" (p. 31). Complicating matters further, increasing state appropriations to higher education alone does not guarantee stable college prices. Unless state support increases more rapidly than campus spending, schools may still feel the need to raise tuition. Indeed, in recent years, public college expenditures have increased rapidly in some states even as public college prices were rising.

On the other hand, New Mexico has successfully expanded their public support of higher education. Beginning in the 1980s, state policy makers undertook a massive increase in state spending on higher education which continued into the early 1990s. This was made easier by the relative health of the state's economy. But rather than reduce taxes or expand other state services, policy makers increased appropriations to higher education by 81 percent between 1984-85 and 1994-95 (New Mexico Commission on Higher Education, 1994).

In most states, however, the prospects of colleges benefiting from additional state tax revenues is bleak. The NEA survey of education committee chairs found that less than 10 percent thought it likely that their state would increase taxes in the next 3 to 5 years (Ruppert, 1996, p. 35). A more likely source of new state revenue for public colleges are bond issues and the proceeds from state lotteries. Bond issues may be proposed to expand or renovate campus facilities. Several states already earmark lottery funds for education and several more are considering the option. One example is the HOPE program in s which uses lottery revenues to pay 100 percent of tuition and fees for all Georgia residents who attend an in-state college and who meet certain other eligibility requirements (Ruppert, 1996, p. 35).

In this difficult fiscal situation, often the best that higher education can hope for, is protect existing funding levels from cuts. In Louisiana, for example, the legislature debated creation of a "floor" for higher education funding. This action would have amended the state's constitution to guarantee that higher education could not receive less funding than it had in the previous year without a two-thirds vote of the legislature. While the measure was not adopted in Louisiana, such earmarking may be a way for state's to stabilize the dramatic shifts which often occur in higher education appropriations (Ruppert, 1996, p. 30).

Changing Pricing Practices

Faced with the realization that there is unlikely to be substantial new nontuition revenue available to them, public college leaders have sought to find ways to maximize their tuition revenues. One way to do this is to change the way they set prices without limiting the access available to lower income and disadvantaged residents. States have tried to do this in a number of ways. The most controversial is to abandon the long held commitment to low tuition.

Questioning the Value of Low Tuition

Historically, state governments and public colleges have shared a common interest in keeping public tuition charges low. States benefited from the economic and social development which accrued from increased participation in higher education, and states saw low tuition as the most direct way to increase the levels of participation in public higher education. Moreover, because colleges are attended disproportionately by higher income residents, a low tuition policy was an easy way to gain the political benefit of distributing public benefits to well-to-do residents.

Public colleges also saw low tuition as beneficial. Low tuition gave them an advantage in the competition with private schools for the best students, and also allowed them to attract large numbers of first generation college students who might otherwise not have attended college. The increase in college participation thought to result from low tuition was seen as fueling economic development and generating new revenues for state governments. More recently, however, as the budgets of both state governments and public colleges have been squeezed by rising costs and falling revenues, the interests of states and public colleges began to diverge. Their different views of how best to respond to this fiscal stress are the natural result of the different constituencies that each serve.

Governors and state legislators focus their attention on the big picture of state finance. They must generate sufficient revenue to provide the services demanded by state tax payers and still keep the state's operating budget in balance. But the dual pressures to expand public services and limit tax increases makes balancing state budgets a political mine field. In order to reconcile these conflicting demands, state policy makers have few options. They can either raise taxes, cut state spending in other areas, increase the efficiency of service delivery, or develop alternative revenue sources. From the perspective of the governor's mansion and the statehouse, higher education looked like a good place to cut.

Public college leaders saw the slowing of state appropriations for higher education as the shortsighted and misguided efforts of state governments to shift the responsibility for funding higher education to students and their families. On campus, this raised concerns that a disastrous chain of events was being set in motion. Rising college costs would lead to lower college participation rates, which would lead to declining economic growth in the state, which would lead to reduced state revenues, which would require further cuts in all types of state services. In the end, students, colleges, and states, would all end up in worse economic condition.

Thus the relationship between states and public colleges in the 1990s is characterized by an unusually high degree of budgetary conflict. Both sides argue that the other does not understand the fundamental problem. But the differences of opinion are not simple misunderstandings. The college affordability problem, and the appropriate response to the problem, look quite different on the green at a public college than in the deliberations of a state appropriations committee (Mumper, 1996).

The High Tuition/High Aid Pricing Model

In order to generate additional revenue without limiting affordability, some states have abandoned their traditional commitment to low tuition and shifted to a high tuition/high aid pricing model. In doing so they shift their funding of higher education away from broad based state appropriations to institutions and towards need-based student aid. The changes forces public college prices to increase. But needy students will be able to pay those rising costs with the larger grants from the state they now receive. On the other hand, higher income students would find their college costs increasing as reduced state support to colleges led to higher tuition. This strategy, which has been followed in Vermont for many years, is now under consideration in several other states as a way of maintaining access for disadvantaged students while dramatically reducing the state's appropriation for higher education.

While the concept of high tuition/high aid funding has been around for more than three decades, it has emerged as a leading reform option in the 1990s. This rise to prominence was driven largely by the fiscal pressures facing state governments. As states searched for ways to reduce spending without producing negative social consequences, high tuition/high aid made sense. Supporters of this pricing strategy argue that low tuition policies are both inefficient and ineffective ways to fund higher education (Wallace 1992). State instructional subsidies allow public colleges to charge students the same low tuition regardless of their need. Students who can afford to pay the full cost of their education pay the same price as those who cannot. Further, low tuition policies insulate public higher education from the discipline of the marketplace (McPherson and Schapiro, 1991). Many students, especially low income students, have little choice but to attend a nearby public college. Thus public colleges do not face the competitive pressures which serve to increase performance and improve quality. They have a captive audience of lower income students who, even with federal financial aid, cannot afford to attend a private college or even another state institution.

Advocates of this approach see it as a way to simultaneously reduce government spending and increase educational quality. By reducing instructional subsidies state governments save money. By increasing student aid, lower income students experience no declines, or perhaps even increases, in college access (Fisher, 1990). And public colleges are forced to raise standards to compete with private colleges based on quality of education and not price.

While there has been a great deal written about the theoretical advantages of the high tuition/high aid model, it was not academic arguments which seem to be driving the shifting pricing strategies of the states. Instead, the change is almost always the product of fiscal stress. In discussing the decision to shift to a form of high tuition/high aid funding in Minnesota, Hearn and Anderson (1995, p. 18) quote a former state representative John Brandl, who observed that "when it came right down to it, we passed the legislation because we just didn't have any money". This view is echoed in the 1996 NEA survey of state education committee chairs found that "few legislators currently support shifting funds from institutional support to student financial aid programs, although many acknowledge that a high tuition/high aid strategy is being carefully examined" (Ruppert, 1996, p. 36).

Criticisms of High Tuition/High Aid Funding

Critics of the High Tuition/High Aid pricing strategy make several arguments against the plan. First, they say that as states allow their public college tuition to rise, lower income and disadvantaged students will conclude that higher education is beyond their financial reach. Well before they can fill out the forms to find out what financial aid they may receive, the "sticker shock" of higher tuition will drive them off of the college track. Others make a financial argument. Because there are more low and middle income students in public colleges, raising tuition for the upper income students will not generate sufficient revenues to cover the financial aid of the lower income students. Moreover, as public college prices rise closer to the level of private colleges, upper income students will be more likely to leave the public system. This will exacerbate the financial stress and force states to further raise tuition and/or reduce financial aid thus undermining the system (Lopez, 1993).

Another set of critics charge that high tuition/high aid pricing generates a political dynamic in state legislatures that will eventually lead to its reversal (Mingle, 1992). The political opposition to tuition increases is strong and broad based. As public college tuition rises, state's feel citizen pressure to moderate those increases. Political support for student aid is weak and narrowly based. Proposals to reduce state student aid meet with little opposition. When state's are under fiscal stress, and looking for ways to reduce spending, the logic of high tuition/high aid pricing asks them to take the heat for tuition increases without any political benefit. As a consequence, even when such plans are enacted, state support for them is likely to erode over time and under fiscal stress (Mumper, 1996). This seems to have been the fate of the approach in Vermont in recent years, where tuition increases have far out paced increases in student aid (Lenth, 1993, p. 36).

Linking Tuition Increases with Student Aid

One way to avoid this dynamic is to explicitly link tuition setting with financial aid. This is the approach taken in Minnesota where an effort was made to more systematically integrate the financing of postsecondary education. A key portion of this plan specified the relative shares of the costs of providing a higher education which would be borne by families and the federal and state government. Students were expected to pay 50 percent of the cost of attendance through income or loans. The remaining costs were to be covered by government assistance. After a student's Pell grant funds were counted, the state of Minnesota accepted the responsibility for any further costs (Hearn and Anderson, 1995).

Similarly, in Washington, the state legislature has taken actions to achieve proportionality between changes in tuition and student aid. In order to do so, they issued a guarantee that if tuition is raised, 24 percent of the increase must go to financial aid (Ruppert, 1996, p. 36-37). This was made easier since the Washington state legislature is responsible for setting tuition levels at the state's public colleges.

Many states have tried different ways to maintain the proportionality between tuition levels and student aid. In Illinois, New York, Rhode Island, and Virginia, the state student aid commissions explicitly attempt to compensate for tuition increases with additional funding for student aid. In Illinois, a rough 'rule of thumb' is employed in which 20 percent of all revenue generated by tuition increases go to student aid. Recently, however, the state has been unable maintain that rule in the face of rapid tuition inflation (Lenth, 1993, p. 24).

Another group of states, including Arizona, North Carolina and Texas, the coordinating or governing board define guidelines within which resources are to be allocated to student aid. But control over administration of the aid programs is decentralized to individual campuses. In Texas, for example, senior public institutions must set aside 15 percent of all tuition income for need based student. Community colleges must set aside 6 percent (Lenth, 1993, p. 25).

A recent study by Carolyn Griswold and Ginger Marine (1996) found that while such linkages do improve the power of the pricing strategy to improve equity within the state, the troubling political dynamic remains. They find that "When aid funding depends on state appropriations, it seems unavoidable that it will be cut during times of financial stress. Similarly, when tuition increases depend on the will of elected officials, such increases will be driven by fiscal considerations" (p.383). They conclude that any attempt to change to a high tuition high aid pricing policy will require extensive planning and coordination as well as strong political will.

Raising Non-Resident Tuition

Another way that public colleges can raise additional revenues is to increase the tuition charged to nonresident students and use that money to subsidize resident students. Today almost every state charges nonresident students a tuition which approaches the full price of the education they receive (Lenth, 1993, p. 16). This is because state policy makers are generally unwilling to provide higher education, or any state service, to those who don't pay taxes in the state. But in both Vermont and Colorado, some public colleges have raised nonresident tuition rates to levels substantially higher than full- cost and use the additional revenue to subsidize in-state students.

Such an approach is unlikely to be successful in all states. Vermont and Colorado are popular destinations for prospective students from all corners of the nation. Public colleges which are located in less desirable places may find that raising tuition charges for nonresidents would reduce their enrollments. This would limit diversity on campus and potentially produce a reduction in campus revenues. Even in Vermont, the state Commission on Higher Education warned that "the rising tuition level charged to out-of-state students threatens the attractiveness of going to college in Vermont". They went on to recommend out-of-state tuition should be "sufficiently attuned to market conditions that a large number of students from outside Vermont are still attracted" (Vermont Higher Education Commission, 1989, p. 5).

RESPONSE #3: Redesigning Delivery Systems

States have also responded to pressures to control college prices by attempting to redesign their higher education system in ways that produce the same output at a lower price. These changes, ranging from comprehensive review of educational programs in order to reduce duplication and eliminate unnecessary offerings to the increased use of instructional technology, are intended to offer citizens the educational services they need at a lower cost to the state budget. While these efforts are quite popular with government leaders and voters, they are generally viewed with skepticism by more traditional higher education institutions (Blumenstyk, 1995).

Distance Learning and Virtual Universities

Perhaps the restructuring proposals that have received the most attention in recent years are those which focus on ways to use new technology to provide higher education to more and different students at a lower cost. These proposals range from two-way interactive television networks where traditional campus courses are brought to previously unserved areas by "virtual campuses" where entire programs are offered through the Internet.

Traditionally, distance learning initiatives have been aimed at nontraditional students and most of it offered through correspondence courses. Recently, however, technology has expanded the focus of distance learning to include more traditional students as well (Institute for Higher Education Policy, 1994, p. 7). To its supporters, these programs are seen as less expensive ways to offer educational services to more students or to handle the pressures of enrollment growth.

Improved technology has been largely responsible for this explosion of interest. The Institute for Higher Education Policy estimates that as recently as 1987, only ten states offered a distance learning program. By 1992, however, all 50 states were operating some type of program (1994, p. 1). The NEA survey of education committee chairs found a remarkable level of agreement on the hope that technology could be used to address a number of higher education problems, especially rising prices. All of the respondents endorsed the expanded use of technology for delivering higher education instruction and 95 percent believe that their legislature will continue to support such programs in the future (Ruppert, 1996, p. 35). James Mingle observes that: "Right now, technology seems to be the only thing states are willing to invest in" (Blumenstyk, 1994).

Perhaps the state which has developed the most elaborate system of distance education is Maine. The Education Network of Maine, a campus-less entity of the University of Maine, now employs an extensive microwave transmission system and a one way video network. The system is now available to over 100 locations and enrolls more than 7,500 students across the state (Blumenstyk, 1996).

West Virginia has also been active in pioneering distance learning. In 1988, West Virginia began a network to deliver one way video and interactive audio to test sites across the state. Today, courses are offered at up to fifty down link sites, including schools, libraries, and hospitals, across the state. The program has served more than 11,000 students and averages 2,100 annually (SHEEO, 1994).

But these programs are dwarfed in comparison to the ambitious plan now being developed by 11 western states. The governors of these states have agreed to explore the creation of a "virtual university" that would deliver courses through computer networks, television, and other technologies, and would award degrees of its own. Unlike traditional institutions that teach and offer credentials, the virtual university would simply award credentials using a set of measures that assess students' mastery of various subjects (Blumenstyk, 1995). In this way, it is similar to New York's Empire State College in offering competence based degrees.

Just how this new university will function is still unclear. But in these states where school-age populations are growing, and funding for new campuses is unlikely to be available, such alternative delivery systems have enormous appeal. Legislators and governing boards see higher education as too labor intensive and are looking to technology to improve productivity and reduce costs. Utah, which anticipates that the number of students attending its public colleges will double by 2010, is banking on such programs to accommodate that growth. Commissioner of Higher Education Cecilia Foxley admits that "you can't handle all of the growth through technology" but she goes on to make clear that Utah "is not planning to replicate nine existing campuses. We will keep building to a minimum" (Blumenstyk, 1994).

Still in spite of its promise, questions remain over whether technology based instruction really reduces costs and thus holds potential for reducing public college tuition. When compared to the expense of building campuses, virtual universities and distance learning plans certainly appear to be less expensive. But the start-up costs of such programs can be substantial. Maine has invested more than $15 million to set up its microwave transmission system. Fiber Network installation costs nearly $10,000 per mile and the setup of each classroom can run another $40,000 (Institute for Higher Education Policy, 1994). Computer and Internet instruction are less expensive, but even here the costs are not insignificant. Moreover, the costs of student advising, electronically available library books, and royalties for copyrighted teaching materials will all increase substantially. As a result, these initiatives may serve to improve access to higher education, but it remains unclear whether they will lower costs as well.

Shifting to Lower Cost Providers

The cost of offering any educational course or program varies widely from institution to institution. If a state is able to shift enrollments from higher cost providers, like flagship research institutions, to lower cost four-year institutions or community colleges, they can reduce the cost of instruction to the student and to the state. In this way, state mandated enrollment caps have long been used to force students to attend less costly institutions. In an effort to shift enrollments to two-year and lower cost four-year colleges, California, Colorado, and Washington have had long standing enrollment caps. Oregon and North Dakota have recently followed suit (Zumeta, 1995, p. 88). While these limits may save money in the short term, they also raise substantial problems. Low income and disadvantaged students are likely to bear the brunt of the these limitations since remaining spaces are usually allocated by raising admission standards. While this would be troubling in any state, it is especially difficult in those states where the number of high school graduates is growing rapidly (Zumeta, 1995).

Differential tuition has also been used to shift enrollment to lower cost institutions. By widening the gap between the price of the flagship research university, where the per student cost of instruction is high, and community colleges, where costs are substantially lower, states can make it more attractive for students to choose the lower cost option. In this way market forces may work as well as state mandates. Sometimes this leads students to enroll in a community college rather than a research university. Other times it means that students will "double dip", taking classes at both institutions simultaneously. At the University of Nevada-Reno, for example, 30 percent of all undergraduates are enrolled concurrently at a community college. The same is true at Arizona State University. At Eastern Michigan University the percentage is about 20 (Gose, 1995).

For such plans to reduce the cost of higher education, however, students must be able to easily transfer those credits from one campus to another. If students are forced to take more courses, even if they are offered at a lower price, they may see their costs actually rise. Consequently, several states have also passed legislation to streamline the articulation and transfer process between two-year and four-year colleges. This is seen as a way of reducing net college prices by allowing more students to study at less expensive community colleges for two years and then transfer to a four year college with no loss of credit.

In Ohio, for example, all public four-year colleges are required to accept any "transfer module" which has been approved by the Board of Regents. In this way any student taking an approved module at a two-year college can be assured that it will transfer to any public four-year college in the state without the loss of credits (Ohio's Managing for the Future Taskforce, 1994).

Response #4: Increasing the Resources of Families and Students

Each of the responses discussed previously seeks to insure that state residents can afford higher education by holding down the prices charged by public colleges. But states have also responded to the affordability problem by developing mechanisms which help families to accumulate the resources necessary to pay those rising costs. Some states have done this by developing innovative mechanisms that allow families to more easily save the money they will need to pay for their children's higher education. By making it easier for families to save, states can shift the responsibility for paying for college from the taxpayers to families and still help those families to pay those higher costs.

State Prepayment and Savings Bond Programs

One way a state can increase the resources available to students and their families is to create state sponsored savings programs. These programs offer benefits to state residents who are willing to save for their children's education. By offering tax-free status to certain investments or a state match of funds placed in certain accounts, states can increase the pool of funds available to families.

These programs fall into two categories; college prepayment programs and college savings bond programs. Tuition prepayment programs allow people to purchase contracts that are guaranteed to cover a percentage of the future costs of attending college. Parents, relatives, and family friends can pay now for a future student's education. College savings bond programs use public relations campaigns and financial incentives to encourage families to save for college by investing in tax exempt government bonds. While neither type of program has much impact on college affordability in the short term, they are designed to change family savings patterns in the long term and make it possible for parents to re-assume a greater responsibility for the college costs of their children (Baum, 1990).

Both types of college savings programs have proliferated since the mid-1980s, at least in part, because they gave state policy makers politically and economically acceptable alternatives to more costly and controversial responses to increasing college prices (Mumper and Anderson, 1996). By creating college savings programs, policy makers could respond to public concerns that college was becoming unaffordable without having to appropriate new spending in the current year.

Michigan was the first state to enact such a program when it established the Michigan Education Trust (MET) in 1986. This program allowed parents and others to pay for the cost of tuition and fees at a state college years before a child reaches college age. The purchase was then guaranteed to cover those costs, no matter how high, when the child eventually enrolled. This program attracted a great deal of attention nationwide. Jeffrey Lehman (1993, p. 28) called the enactment of MET "the most widely publicized government action in the field of higher education finance during the 1980s". Soon programs like MET were under consideration everywhere. A recent survey by the General Accounting Office found that more than 40 states had considered enactment of some type of savings program and 30 of those had actually adopted some type of program. Of those, 12 have adopted a prepayment program and 7 of those have fully implemented that program (GAO, 1995).

The operation of a state prepayment program is a fairly complex matter[5]. The revenue from purchaser payments are pooled into one large fund and invested with the goal of achieving a rate of return that exceeds the inflation rate at participating institutions. Each semester that the beneficiary enrolls in a participating college, the program pays the school whatever it currently charges for tuition and fees and any other prepaid benefits. If the prepaid benefits are not used as intended, the funds are then refunded according to program stipulations.

In 1995, Massachusetts introduced a new type of prepayment plan. This program sells "tuition certificates" redeemable toward the cost of tuition and fees at any of the 67 public and private colleges in the state that agreed to participate in the program. The certificates are guaranteed by the state to hold their value until redeemed by the beneficiary. For example, if a $1,200 certificate is equal to 20 percent of tuition costs at a given college at the time of purchase, the certificate will cover that same percentage of costs when the beneficiary enrolls in that college in the future (Healy, 1995).

State prepayment programs have a number of limitations which have reduced their appeal to states in recent years. Most participants in these programs come from middle and upper income families. The GAO (1995) reported that:

> In Alabama, Florida, and Ohio, the majority of purchasers reported family incomes of over $50,000 in 1992, while the majority of state families with children had incomes under $30,000. In addition, Alabama state tax returns from 1992 and 1993 revealed that the median income among purchasers was about $61,200 while the Bureau of the Census data showed that the 1992 median family income for all families in the state was about $27,400 (p. 5).

Because of these participation patterns, the programs could subsidize their mostly well-off participants while doing little to help lower income families.

Another major concern over prepaid tuition programs is the degree of risk they pose for states. Critics worry that they could create an unfunded liability for the state if investment income is insufficient to cover the programs obligations (Lehman, 1995, p. 30-31). However, no state program has yet experienced such a

[5]Details of the development and operation of these programs are explained in Horvitz, 1993.

shortfall and, if one did occur, it is unclear how the situation would be resolved. Because of these concerns, more states have chosen to establish a college savings bond program rather than a prepaid tuition program. Issuing college savings bonds are both less financially risky and easier to administer. About 20 states have sold college savings bonds, though relatively few have done so on a regular basis (GAO, 1995).

State college savings bond programs are quite straightforward. The state issues general obligation, zero-coupon bonds, and markets them to individuals wanting to saving for future educational costs. Because these bonds are state debt instruments, the interest earned is exempt from state and federal taxes. Although these bonds are marketed as college savings bonds, in practice they generally do not require the purchaser to spend the funds on higher education and the purchaser need not designate a beneficiary when the bond is purchased (Williams, 1993).

Finally, Kentucky has developed a college savings account program. People can save as much or as little as they like on behalf of a designated beneficiary, depending on their individual savings goals. Deposits may be as low as $25. The program guarantees a minimum 4-percent rate of return and the interest is exempt from state income taxes. When withdrawn, the funds can be spent at virtually any college in the country (GAO, 1995, p. 6).

Tuition Waivers and Exemptions

Many states offer some full or partial tuition waivers for certain types of students. This is quite different from need-based student aid in that the waivers are awarded based on some criteria other than financial need. Louisiana, North Dakota, Virginia, Wisconsin, and Wyoming among others, all provide tuition waivers to top high school students who attend their states public colleges. In Illinois, state statutes provide each member of the state legislature with two tuition waivers each year to be used at the discretion of the member at an Illinois public college or university (Lenth, 1993, p. 22-23). But it is Texas which offers the most extensive system of tuition waivers and exemptions. In 1993, nearly 23,000 students at Texas colleges and universities received tuition and fee exemptions totaling $9 million. These waivers go to war veterans and veterans who served at least 180 days during the cold war, children of disabled firefighters and peace officers, and the highest ranking graduate of each Texas high school (Texas House Research Organization, 1995).

These waivers and exemptions may serve an important social purpose by rewarding certain behaviors and life choices. But they do little to improve college affordability for those who are not part of the narrow group that is eligible for benefits. Indeed, by reducing the revenue flowing into public colleges, these waivers may actually cause some colleges to raise their tuition for other students in order to compensate for the lost revenue.

Evaluating the State Responses

There are a great many studies and reports which describe or analyze the efforts of particular states to keep their public colleges affordable. There are also many studies which examine the operation of a single reform strategy among a number of states. But there has been almost no systematic evaluation of the wide range of efforts which have occurred across the 50 states.

The only major study which attempts to explain the variations in public college affordability among all the states was conducted by Hearn, Griswold, and Marine (1996). Using 1990 data, they disaggregate the impact of several independent variables on public college tuition and student aid levels in the states. Their study makes several important findings. Region of the country is the variable which they identify as most closely associated with tuition and student aid in the states. States in the northeast and midwest have high tuition and high aid levels. States in the southwest and west have low levels of tuition and aid. Second, they find that economically developed states are more likely to have low-priced entry points into the postsecondary system. The higher the personal income in a state, the lower the tuition at two-year colleges. Finally, they found that states with planning agencies and strong coordinating boards were associated with higher tuition levels at four-year colleges (p. 267-269).

These findings point to important underlying factors that may drive changes in college affordability. The more fully the complex dynamics between tuition, student aid, state politics, and governance structure are understood, the easier it will be for policy makers to develop paths to productive policy change. But, at least in the short term, such findings provide limited guidance for policy makers. A state's region and its level of economic development are largely fixed. Even the governance structure of the state's colleges may be quite difficult to adjust.

There are still no studies which systematically examine the link between *specific policies* and college affordability across the states. In spite of the absence of such empirical analysis, however, it is clear that no policy, program or approach has fully addressed the problem. It persists, to one degree or another, in every state. This is not to say that the problem is unsolvable. Many of the responses discussed here have been in place for only a short time. Moreover, since the precise causes of the problem may vary considerably from state to state, it is perhaps unreasonable to expect to find a single solution that works everywhere. Still, with these limitations in mind, it is possible to draw some speculative regarding the effectiveness of these state efforts.

Paths to Improved Affordability

There are two sharply different paths that seem to have produced at least modest success in maintaining public college affordability. The first is for states to insure that their public colleges hold the line on tuition increases. A state can do this through generous support of its public colleges, or it can be done through threats and mandates to the colleges. This low tuition strategy, most evident in Texas, North Carolina, Idaho, Nevada, and Arizona, is generally a very costly approach. But it has had a long record of success in many states in the 1960s and 1970s. Low tuition reduces the need for states to provide financial aid to students or devise incentive programs to encourage families to save for college. This, in turn allows these states to direct all of their support for higher education to be directed to the single purpose of low tuition.

In addition to its possibilities, however, a low tuition strategy also involves substantial risks. If low tuition is achieved by starving a state's colleges of the funds necessary to maintain high quality programs or to attract the best faculty and students, it may prove a hollow victory. Policy makers must carefully answer the question "access to what?" when considering this approach. Similarly, if low tuition is achieved by limiting enrollments at some or all institutions, it may prove counter productive. Lower income and disadvantaged students may be denied access to a state's best colleges, not because they cannot afford it, but because all the slots are already filled.

Finally, even those states which have been able to maintain the lowest tuition at their public colleges have still experienced substantial price increases in the past few years. The average public college tuition in every state increased by more than 75 percent between 1985 and 1995. The average tuition doubled in all but five states. This is far greater than the increases in consumer inflation or family wages during the same time period. As prices begin to creep up, even the lowest price public colleges risk becoming unaffordable. This is because students do not only pay tuition when they go to school, they must also pay living expenses, buy books, and get back and forth to classes. Additionally, students must pay an opportunity cost to enroll in college. They must forgo earnings and job experience while they are in school. Unless there is student aid, family savings, or loans to offset these nontuition costs, even low tuition colleges may prove too costly and many potential students may simply decide not to enroll.

During the last two decades, state after state has abandoned this low tuition approach either because they could no longer afford its high cost, or because they had lost confidence in its effectiveness. California, New York, and Washington are recent examples. In its place, states have begun to employ a combination of steps that attack the problem from several sides at once. While the precise combination of steps varies from state to state, they usually include efforts to improve efficiency and productivity, to generate new revenues, and to protect the most vulnerable students from the impact of tuition increases. This incremental approach has also produced some success in maintaining public college affordability.

In most states, the first step in this approach has been to search for ways to offer the same or improved educational services at a lower price. This can be done through a redesigned delivery system, efficiency improvements, workload increases, and/or shifting students to lower-cost providers. It can also be done through performance funding. While these efforts alone are unlikely to solve the problem, over time they can reduce its magnitude. They can also help to restore the confidence of policy makers and the public that public college leaders can be trusted to spend their funds wisely.

The second step in this incremental strategy is to expect more affluent students to pay closer to the full price of their education. This will mean higher tuition for many students which will, in turn, generate new revenues for the state's public colleges. These new tuition revenues can then be used to fund need-based student aid programs or simply to replace those dollars which are no longer supplied by the state government.

The final step is then to insure that the most disadvantaged students are protected from the impact of rising prices. In most cases, this has been done by expanding the availability of need based financial aid. This can involve a full-scale shift to high tuition/high aid funding, or simply a more careful coordination between the levels of need-based aid and the rates of tuition increase. Regardless of how it is done, no effort to improve college access can be effective if it does not insure that the most disadvantaged students have access to the resources necessary to participate in the system. Some states have done this by tracking many such students into colleges with lower instructional costs. This can be done by a system of differential tuition where community colleges are priced much lower than research universities. It can also be done by a system of enrollment caps that close many of those students out of the most high cost institutions.

Like the low tuition strategy, this incremental approach also carries substantial risks. It is difficult for anyone to oppose improved efficiency and productivity. But what is an unnecessary or wasteful program to one person is often critically important to another. As such, the price of state efforts to improve efficiency may be discord and acrimony which can disrupt the smooth functioning of campuses and programs.

At the same time, rising tuition can undermine the political and public support for higher education funding. If more affluent families begin to abandon a state's public college in favor of private or out-of-state institutions, governors and legislators may lose interest in the need to maintain affordability. Similarly, during times of economic downturn, states may slash their financial aid programs because they are not protected by strong political interests. And too generous state savings programs may pull limited subsidy dollars away from programs which support all students to pay for benefits that go primarily to the richest families.

In the end, neither the low tuition nor the incremental approach represents a final solution. The myriad of social, political, and economic problems that have caused the fiscal problems now faced by state governments seem to be here to stay. As a consequence, no state is likely to have the resources necessary to turn the clock back to the era of no, or very low, tuition. As public colleges try to cover their increasing expenditure levels with more and more limited state dollars, they will be forced to raise tuition even further. Without sufficient funds to compensate for those increases with proportionate increases in need-based student aid, public college affordability will continue to decline. As long as these fundamental conditions remain in place, nothing appears likely to reverse the present trends. The best that states and institutions can hope to do is moderate the rate of declining affordability and to protect the most disadvantaged from its full effect.

References

Baum, S. (1990). The need for college savings. In J. Hansen (ed.) *College Savings Plans: Public Policy Choices*. New York: The College Board.

Berdahl, R., and McConnell, T.R. (1994). Autonomy and accountability: some fundamental issues. In P. Altbach, R. Berdahl, and P Gumport (eds.) *Higher Education in American Society* (3rd ed.). Amherst NY: Prometheus Books.

Bergmann, B. (1991). Bloated administration, bloated campuses. *Academe* 77 (November/December): 12.

Blumenstyk, G. (1994). Networks to the rescue? *Chronicle of Higher Education*, December 4, 1994. A21.

Blumenstyk, G. (1995). Campuses in cyberspace. *Chronicle of Higher Education*, December 15, 1995. A18.

Blumenstyk, G. (1996). Learning from afar. *Chronicle of Higher Education*, May 31, 1996. A15.

Cage, M. (1995). Regulating faculty workloads. *Chronicle of Higher Education*, January 20, 1995. A31.

Chronicle of Higher Education. (1994). Almanac issue. September 14, 1994.

Colorado Commission on Higher Education. (1994). *Access to Success: Accommodating Student Enrollment Demand For Colorado Higher Education*. Denver: Colorado Commission on Higher Education.

Fisher, F. (1990). State financing of higher education: a new look at an old problem. *Change* 22: 42-56.

Francis, C. (1990). *What Factors Affect College Tuition?* Washington, DC: American Association of State Colleges and Universities.

Gallup Organization. (1991). *Attitudes About American Colleges*. Washington, DC: Council for the Advancement and Support of Higher Education.

General Accounting Office. (1995). *College Savings: Information on State Tuition Prepayment Programs*. Washington, DC: General Accounting Office.

Gladieux, L. and A. Hauptman. (1995). *The College Aid Quandary*. Washington, DC: Brookings Institution.

Gose, B. (1995). Double dippers. *Chronicle of Higher Education*, August 4, 1995. A27.

Gold, S. (1995). *The Fiscal Crisis of the States*. Washington, DC: Georgetown University Press.

Griswold C., and Marine, G. (1996). Political influences on state policy: higher-tuition, higher-aid, and the real world. *Review of Higher Education* 19: 361-390.

Hauptman, A. (1990). *The College Tuition Spiral*. Washington, DC: The College Board and the American Council on Education.

Healy, P. (1995). Massachusetts Starts a Novel Program to Help Families Pay Tuition *Chronicle of Higher Education*. February 3, 1995. A 24.

Hearn, J., and Anderson, M. (1995). The Minnesota financing experiment. In E. St. John (ed.) *Rethinking Tuition and Student Aid Strategies*. San Francisco: Jossey Bass.

Hearn J., Griswold, C., and G. Marine, G. (1996). Region, resources, and reason: a contextual analysis of state tuition and student aid policies *Research in Higher Education* 37: 241-278.

Hines, E. (1996). *State Higher Education Appropriations 1995-96*. Denver: State Higher

Education Executive Officers.

Horvitz, P. (1993). Prepaid tuition plans: an exercise in finance, psychology, and politics. In M. Olivas (ed.) *Prepaid College Tuition Plans: Promise and Problems.* New York: The College Board.

Hsu, S. (1995) Va. failing in funding, colleges say: business leaders warn of cuts' effect *Washington Post.* January 23, 1995. A1.

Institute for Higher Education Policy. (1994). Assessing distance learning from a public policy perspective. *Policy Steps* 3:7-9.

Layzell, D. (1996). Developments in state funding for higher education. In J. Smart (ed.) *Higher Education: Handbook of Theory and Research,* Vol. XI. New York: Agathon Press.

Lehman J. (1993). The distribution of benefits from prepaid tuition programs: new empirical evidence about the effects of program design on participant demographics. In M. Olivas (ed.) *Prepaid College Tuition Plans: Promise and Problems.* New York: The College Board.

Lenth, C. (1993). *The Tuition Dilemma—State Policies and Practices in Pricing Public Higher Education.* Denver: State Higher Education Executive Officers.

Lively K. (1995). Tough state budget line. *Chronicle of Higher Education.* January 20, 1995. A23.

Lopez, M. (1993), High tuition high aid won't work. *Chronicle of Higher Education.* April 7, 1993. B1.

Magner, D. (1995). Restructuring stirs outcry at James Madison. *Chronicle of Higher EducationI.* March 3, 1995. A15.

Massy, W., and R. Zemsky, R. (1990). *The Dynamics of Educational Productivity.* Denver: State Higher Education Executive Officers.

Mathesian, C. (1995). Higher ed.: the no longer sacred cow. *Governing.* March: 20-24.

McGuinness, A. (1994). The states and higher education. In P. Altbach, R. Berdahl, and P. Gumport (eds.) *Higher Education in American Society* (3rd ed.). Amherst NY: Prometheus Books.

McPherson, M., and Schapiro, M. (1991). *Keeping College Affordable.* Washington, DC: Brookings Institution.

Miller, V. (1993). State medicaid expansion in the early 1990s: program growth in a period of fiscal stress. In D. Rowland et. al. (eds.) *Medicaid Funding Crisis.* Washington, DC: American Association for the Advancement of Science.

Mingle, J. (1992). Low tuition, progressive taxation. *AGB Reports.* Washington, DC: Association of Governing Boards.

Mortenson, T. (1994). retrenchment in allocation of state resources to public higher education paused in FY 1995 *Postsecondary Education Opportunity.* November: 7.

Mumper, M. (1993). Maintaining public college affordability in the 1980s: how did the states do? *Journal of Education Finance.*19: 183-199.

Mumper, M. (1996). *Removing College Price Barriers.* Albany NY: State University of New York Press.

Mumper, M., and Anderson, J. (1996). Helping Families Save For College. In M. Mumper. *Removing College Price Barriers.* Albany NY: State University of New York Press. 175-190.

Nedwek, B. (1996). Public policy and public trust: the use and misuse of performance indicators in higher education. In J. Smart (ed.) *Higher Education: Handbook of Theory and Practice,*Vol. XI. New York: Agathon Press. 47-89.

New Mexico Commission on Higher Education. (1994). *The Condition of Higher Education in New Mexico.* Santa Fe, NM: New Mexico Commission on Higher Education.

Ohio Governor's Task Force. (1992). *Managing for the Future: Challenges and Opportunities for Higher Education in Ohio.* Columbus OH: Ohio Board of Regents.

Ruppert, S. (1996). *The Politics of Remedy: State Legislative Views on Higher Education.* Washington, DC: National Education Association

Schmidt, P. (1996). Earning appropriations *Chronicle of Higher Education.* May 24. A23.

SHEEO. (1994). *Access Through Distance Education: Collaborative Ventures in West Virginia.* Denver: State Higher Education Executive Officers.

Sowell, T. (1992) The scandal of college tuition. *Commentary* 95:24.

Texas House Research Organization. (1995). *State Finance Report Number 74-2. Higher Education Tuition in Texas.* Austin, TX: Texas House of Representatives.

Tucker, J., and Voelker, J. (1995). Public higher education policy in Ohio. In C. Lieberman (ed.), *Government, Politics, and Public Policy in Ohio.* Akron, OH: Midwest Press.

U.S. Department of Education. (1995). *Digest of Education Statistics.* Washington, DC: U.S. Government Printing Office.

Vermont Higher Education Study Commission. (1989). *Recommendations and Findings.* Montpelier, VT: Vermont Higher Education Study Commission.

Virginia Council on Higher Education. (1994). *Restructuring Criteria.* Charlottesville: Virginia Council on Higher Education.

Virginia Governor's Commission on Government Reform. (1994). *Executive Committee Briefing.* Charlottesville: Governor's Commission on Government Reform.

Wallace, T. (1992). The inequities of low tuition. In *Tuition and Finance Issues for Public Institutions.* (AGB Occasional Paper #15), Washington, DC: Association of Governing Boards.

Williams, D. (1993). Taxation on prepaid tuition and other forms of college expense assistance. In M.A. Olivas (ed.) *Prepaid College Tuition Plans: Promise and Problems.* New York: The College Board. 55-77.

Zumeta, W. (1995). State policy and budget developments. *The 1995 NEA Almanac of Higher Education.* Washington, DC: National Education Association.

Zumeta W., and Looney, J. (1993). State Policy and budget developments in higher education. Working paper in public policy analysis and management. Graduate School of Public Affairs, University of Washington, Seattle WA

Discriminant Analysis in Higher Education Research

Carl J Huberty and Laureen L. Lowman
University of Georgia

1. INTRODUCTION

Most empirical research in higher education involves the study of multiple characteristics of students, faculty, and institutions. Collections of such analysis units may be in a single group or in multiple groups. Purposes behind the construction of multiple-variable data sets may involve the study of prediction, identification, group comparison, typology discovery, relationship, characteristic structure, data reduction, and causal modeling, as well as other research purposes. That is, what is very often of interest in higher education research is the study of multiple variable measures which would call for a *multivariate analysis*. [See Huberty (1994a and 1994b, pp. 32-34) for further discussion on multivariate analyses in general.] The spectrum of multivariate analysis includes multiple regression/correlation analysis, multivariate analysis of variance, cluster analysis, multidimensional scaling, canonical correlation analysis, factor analysis, pattern recognition, principal component analysis, path analysis, and discriminant analysis. It is the latter that is the focus of this chapter.

A scanning of books and articles written by statisticians, methodologists, and empirical researchers leads one to conclude that "discriminant analysis" means different things to different people. For some, it involves predicting membership in groups on the basis of multiple analysis unit (or subject or person or object) attributes. For example, it may be of interest to study the predictability of college/university faculty rank using predictor measures such as educational level, years since degree, average merit score, and previous employment position. For others, discriminant analysis involves the description of group differences or of grouping variable effects. For an example of this type of study, suppose that it is of interest to compare universities with respect to, say, six academic environments and five institutional environments in terms of a number of undergraduate teaching goals. That is, an intent of such a study is to assess the effects (singly and jointly) of academic environment and institutional environment on undergraduate teaching

J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*

goals. With either of the above views of discriminant analysis, there are two or more groups of analysis units and a set of response variables; that is, there is a set of unit characteristics or attributes in addition to the grouping variable(s).

There are, then, two reasons for conducting a discriminant analysis:

- Prediction of group membership
- Description of grouping variable effects

With the former, the multiple response variables play the role of predictor variables, whereas with the latter, the multiple response variables play the role of outcome variables. The analysis for the former may be termed a *predictive discriminant analysis* (PDA), whereas for the latter it is *descriptive discriminant analysis* (DDA). For the moment, let G denote a grouping variable, and $\underline{V}$ denote a collection (vector) of response variables. Schematically, we may represent the research purposes of the two types of discriminant analyses for a one-factor design as:

PDA: $\underline{V} \rightarrow G$

DDA: $G \rightarrow \underline{V}$

It should be noted that, analysiswise, the DDA arrow may be bidirectional; that is, DDA techniques may be viewed as correlational in nature. More on the distinction between PDA and DDA is given by Huberty (1994b, pp. 26-31), Huberty and Barton (1989), and in the fifth section of this chapter.

The next two sections of this chapter are devoted to applications and illustrative discriminant analyses related to higher education. First a PDA will be discussed and illustrated, followed by a DDA. Real data (obtained via a faculty survey) will be used to illustrate the analyses. The survey was conducted in 1989 by The Carnegie Foundation for the Advancement of Teaching. Survey items pertained to respondent current academic position, scholarly activities, home institution, home institution undergraduate students, retirement plans, and demographics; as well as tenure, college curriculum, and higher education in the United States.

The fourth section of this chapter pertains to the use of discriminant analysis as a supplement to cluster analysis, and the fifth section reviews some issues and problems in discriminant analysis. The chapter is concluded with a comments section.

2. PREDICTIVE DISCRIMINANT ANALYSIS IN HIGHER EDUCATION RESEARCH

Exemplary PDA Research Questions

Let us begin by considering some recently published empirical studies of interest to higher education professionals wherein PDA was employed. Remember, the basic purpose of such a study would be that of *prediction* of group membership using data on a set of predictor variables. Some examples of research questions, taken from higher education journal articles, are given below:

- How well can faculty rank be predicted using demographics, prior employment, and some "merit" variables?
- To what extent can institutions be identified with higher education taxonomies on the basis of psychological climate?
- How useful is a predictive model for college student retention that is based on student work values and self-report of competencies?
- How well can sorority rush status of college women be predicted using such variables as family income, frequency of alcohol consumption, college performance, and social self-esteem?
- Using personality measures (obtained via a personal preference instrument), how well can college students be identified with their race/ethnicity/gender groups?

A design of a study in which PDA would be used is, simply, a one-way layout. There is a single grouping variable (or "factor") with two or more levels; that is, there are two or more groups of analysis units (e.g., students, faculty, institutions). On each unit there would be measures on one or more predictor variables. A common practice is to let k denote the number of groups and let p denote the number of predictors.

Some PDA Preliminaries

A Research Situation

To illustrate how a PDA might be carried out in the context of a higher education research situation, an example problem was contrived using some of the 1989 Carnegie Foundation survey information. One research question might be stated as: During the past two years, have you ever considered a permanent departure from academia? Three groups of faculty members may be defined as:

Group 1: I have given it serious consideration

Group 2: I have considered it, but not seriously

Group 3: I have not considered it

A few comments about this grouping variable follow. In a given institution, the definition of these groups may not be as clear-cut as would be desirable. That is, whether a given faculty member is clearly a member of one group or another may depend upon a number of factors that may change over a short period of time within a given institution. For example, group membership may depend upon a faculty member's health or family situation, or upon administrators, or upon changing institution faculty benefit packages. Any faculty member to whom the research question might apply must belong to only one of the three groups if the following related research question might also apply: How well can a faculty member be predicted to be associated with one of the three defined groups?

PDA Variables

So, first of all, for a PDA to be meaningfully applicable, the grouping variable—

which in this situation plays the role of a criterion variable—must be well defined, and every unit of analysis (e.g., a faculty member) that/who might be considered for group-membership prediction must be a member of one and only one of the criterion groups.

Secondly, how measures on the predictor variables are obtained should be clearly described. With the current exemplary research situation, there are 21 potential predictor variables—see Table 1 for a listing of the 21 predictors and how each is indicated and measured. It may be noted that each of the 21 predictors was measured by a single survey item. It is recognized that although this practice is fairly common, it may be troublesome for some higher education researchers, particularly those whose interests focus on measurement. An alternative is to use scores on item composites (based on substantive judgment) as variable scores; this could enhance the reliability of the variable scores. [Composite scores will be used in the DDA illustration later in this chapter.]

Suppose in a general research situation that the grouping variable is defined and an appropriate collection of predictor variables is chosen. Suppose, further, that the collection of predictor variables contains some continuous variables with obviously or naturally underlying continuous measurement scales. In higher education research, some categorical predictors would undoubtedly also be included in the predictor collection, particularly in a survey context. For a two-category predictor (e.g., gender), a 0-1 scale is typically used. For a multiple-ordered-category predictor, *integer scaling* may be used—e.g., academic rank where measures of 1 to 4 are commonly used. For multiple-unordered-category predictors, scaling becomes a bit more involved. [Use of dummy or indicator variables is *not* recommended.] The goal is to develop, using the information on hand, a scale of measurement that underlies the unordered categories. One approach that has some promise for use in any multivariate context is a transformation that was independently developed by R.A. Fisher and H.O. Lancaster—this scaling method is described by Huberty (1994b, pp. 153-154) along with which is provided a diskette containing a Fisher-Lancaster (F-L) computer program.

Another comment about variables used in discriminant analysis in general, and PDA in particular, may apply to some research situations in higher education. As mentioned above, data for higher education empirical studies are often obtained via surveys. Suppose a survey instrument designed for a particular research question involved a fairly large number of items, say 50 items. To claim that the 50 items reliably measure 50 variables may be stretching things a bit. So, to get a better handle on what characteristics the items are more reflectively measuring, some type of "data reduction" analysis may be appropriate. An example of such an analysis is a principal component analysis (see, e.g., Cliff, 1987, chap. 13; Rencher, 1995, chap. 12). Then the response "variables" to be considered are those represented by the (expectantly radically less than 50) p components to be retained for measurement purposes. The response variable measures to be used in the discriminant analysis, then, are the p component scores.

Data Checking

Before proceeding with the discussion on PDA further, a suggestion will now be made. It is strongly advised that prior to conducting a PDA, an inspection of the data matrix on hand be conducted. Such an inspection would involve identifying any missing data and any potential outlying predictor variable score vectors. Mention of the inspection should be made in the final write-up of the research report.

PDA Purposes

Depending upon the goals of the researcher, one or two primary purposes of conducting a PDA may be appropriate. One primary purpose is to assess the accuracy of predicting group membership using the data on hand. The related research question is: How well can group membership be predicted using this set of predictors? Alternatively stated, the intent of such a study would be to assess the predictive power of a set of predictor variables. A side-question in assessing predictive power is: Are there any peculiarities in the prediction results that pertain to individual analysis units?

The second primary purpose of conducting a PDA is to develop a prediction/classification rule (using the data on hand) that may be used with "new" analysis units (e.g., students, faculty). This purpose, one that is rarely apparent in published manuscripts in which PDA is used, is one that is of a very practical, applied nature. [The reader may want to refer to the fifth section of this chapter for a discussion on classifying new units.] For the exemplary research situation introduced above, both research purposes and the side-question of the first will be addressed in the illustrative example presented later.

A secondary purpose of conducting a PDA is to determine a ranking of the predictor variables in terms of their relative contribution to predictive accuracy. As will be illustrated, predictive accuracy used to assess accuracy may pertain to all k groups, or to some particular group(s).

Developing a Prediction Rule

Classification Functions

As some readers know, the form of the prediction rule utilized in linear multiple regression analysis (MRA) is that of a linear composite/combination of the predictor variables. A *set* of composites of the predictor variables is one form of a prediction rule in PDA. In PDA, these composites are labeled *classification functions*. For a k-group design, there will be k such functions. A differentiation, however, between the composites in MRA and those in PDA is that in the latter the composites may be linear or quadratic, whereas in linear MRA the composites are linear in form.

To make a fairly long story short (see Huberty, 1994b, pp. 58-61), a brief discussion is now given on linear versus quadratic composites in PDA. Associated with each criterion group is an array of numbers that reflect, in a way, the variability shared between pairs of variables; these numbers are called covariances.

That is, a covariance for two variables is a reflection of the covariation between the two variables. For each group we have a p x p matrix of covariances (with the p variances on the main diagonal of the matrix). If the k covariance matrices are approximately equal, then the set of k PDA composites to be employed are *linear classification functions* (LCFs); whereas if the k variance matrices are not "in the same ballpark," then the set of k PDA composites that may be employed are *quadratic classification functions* (QCFs). To make an assessment of the equality of the k covariance matrices, a statistical test may be employed. Because such a test is extremely powerful in a statistical sense, it is suggested that the researcher require a very small P value (at least as small as .005) to conclude that the corresponding population matrices are different. [A more extensive discussion of a statistical test of covariance matrix equality and of the linear-quadratic issue is given later in this section.] One form of a PDA group-membership prediction rule, then, is a set of k (the number of criterion groups) predictor composites, either linear or quadratic. An analysis unit is assigned to the group with which the largest composite score is associated.

Prior Probabilities

The (unconditional) probability of any unit belonging to a group is termed a *prior probability.* It is "prior" in the sense that it is not dependent (or conditioned) on any observed predictor scores. These probabilities should reflect the proportional sizes of the k populations corresponding to the k criterion groups. Estimated prior probabilities are specified for the computer analyses by the researcher. So, to conduct a PDA it behooves the researcher to be knowledgeable about the relative sizes of the corresponding populations of interest. Group sizes themselves would be a basis for prior probability estimates *only* if a proportional sampling design was employed. Note that the sum of the k priors is necessarily 1.0. It should be noted, too, that unlike (estimated) posterior probabilities, there are only k (estimated) prior probabilities, period; that is, there are not k priors for each analysis unit.

In conducting a PDA via the use of a computer package, user-specified prior probability estimates are "automatically" incorporated into the prediction rule used. Basing priors on group sizes is the default for SPSS for Windows; to use other priors (which is desirable unless a proportional sampling plan is utilized), one can paste the syntax into a syntax window and edit the command file. It should be noted that if the purpose of PDA is to derive a set of k composites to be used with new analysis units, the respective prior probability estimates need to be incorporated into the composites.

Posterior Probabilities

A prediction rule may also be based on (estimated) probabilities of group membership. For a given analysis unit, these may be described as probabilities of group membership, given the unit's score vector—a *conditional* probability, if you will. Because these probabilities are conditioned on the predictor scores (i.e., *after* the scores are obtained), they are termed *posterior probabilities.* For k

groups, there are k such probabilities for each unit. A unit is assigned to that group with which the largest (estimated) posterior probability is associated. The k posterior probabilities sum to 1.0—this is because it is assumed at the outset that all units of interest belong to one (and only one) of the k groups. Specific definitions of these probabilities are discussed by Huberty (1994b, pp. 45-49, 56-58). It should be noted that two computer software packages (BMDP and SAS) compute posterior probabilities both when a linear rule is specified and when a quadratic rule is specified; the SPSS package correctly computes the posterior probabilities only for a linear rule.[1] A prediction rule to be used with new units may be developed using posterior probability formulation (or by using a set of k predictor composites).

It turns out that either form of a prediction rule—a set of k composites or k (estimated) posterior probabilities—may be used for the second primary purpose for conducting a PDA; namely, to assess the accuracy of predicting group membership. This statement of equivalence assumes that the same information is used with the two forms. One type of information has to do with the covariance matrices—pooled or separate-group—used in the computations to obtain the composite weights and to obtain the (estimated) posterior probabilities. Another type of information has to do with *prior probabilities* (discussed above).

Rule Basis

It has not been mentioned in the discussion in this section, but a prediction rule is most often based on the condition of multivariate normality of the predictor variable collection. For example, the particular predictor composites mentioned above are normal-based. Although estimation of the posterior probabilities of group membership may be based on any distributional form of the predictor collection, without a doubt the most popular basis is multivariate normality.

Other rule bases are the logistic distribution, multinomial distribution, and a "nearest-neighbor" notion. In addition to the use of categorical scaling and Fisher-Lancaster scaling mentioned earlier in this chapter, some work has been done on using rank transformations in connection with PDA. All of the optional rule bases will not be considered in this chapter. Only normal-based rules are discussed here. See Huberty (1994b) for various references to writings dealing with some alternatives.

Misclassification Costs

Suppose one has a research situation that involves classification of analysis units

[1]The computation of posterior probabilities by SPSS is different from that used by SAS and BMDP. Instead of using predictor variable scores as input, SPSS uses scores on linear composites of the predictor variables—these composites are linear discriminant functions (LDFs). The SPSS quadratic rule, then, is one that is quadratic in terms of the LDFs and *not* in terms of the predictor variables themselves. Thus, SPSS "quadratic" results are different (and usually incorrect) relative to SAS and BMDP quadratic results.

into one of two groups. It may be determined prior to the analysis that it is a more serious error of classification if a unit who is truly from group 1 is assigned to group 2 than if a unit who is truly from group 2 is assigned to group 1. If so, these (relative) costs of misclassification should be incorporated into the classification rule. As might be concluded, the incorporation of unequal misclassification costs into a rule applicable to a three- or four-group situation can get quite messy. A brief discussion of misclassification costs is given by Huberty (1994, pp. 68, 136); McLachlan (1992, pp. 7-9) provides a bit more detail.

Predictor Selection in PDA

It is well known that in multiple regression analysis (MRA), the value of R^2—which reflects the (squared) correlation between the criterion variable scores and scores on the "best" composite of the predictor variables—cannot decrease if the number of predictor variables increases. Stated in a reverse fashion, it is less well known that an estimator less biased than R^2 can actually increase with a decrease in the number of predictors. It is this second situation that is quite common in a PDA context. That is, predictive accuracy (i.e., sample hit rates) can very well increase with a decrease in the number of predictors. Thus, the researcher should consider the deletion of predictors in developing the rule to consider for final use (in assessing predictive power or to use with new units).

A very typical procedure used in predictor selection in MRA situations and even in PDA situations is a stepwise analysis. The lack of appropriateness of using popular stepwise discriminant analysis programs for predictor selection is reviewed by Huberty (1989) and Huberty (1994b, pp. 118-122). The selection procedure advocated here is an all-possible subset analysis. On the surface 12 predictors would call for $2^{12} - 1 = 4095$ analyses, including the 12-variable analysis. Oftentimes, this total number of analyses can be drastically reduced by considering some substantively related notions. Based on previous research, expert advice, and common sense, there may very well be a subset of predictors that should be retained for the final rule. If so, this could drastically reduce the total number of analyses to conduct. For example, suppose that in a 12-predictor situation, there are five predictors to be retained no matter what. Then the number of analyses reduces to 127 analyses; this is manageable even without a specific all-possible-subsets computer program.

Once all subset analyses are completed, it takes a series of judgment calls to decide on the subset size and the actual subset to be finally considered. Both of these decisions would be based on hit rate estimates. An example of the judgment process is illustrated later in this chapter.

If a predictor subset is to be considered for final interpretation purposes, the subset should be selected at the *outset* of the study analysis process.

Variable Ordering in PDA

In the conduct of a PDA, it is "natural" for researchers to ask: Which predictors are most important? Or, which predictors contribute most to the predictive accuracy (for a particular criterion group, or across all groups)? Or, what is a rank-ordering of the predictors in terms of contribution to predictive accuracy? It is natural, too, for many applied researchers to conduct a "stepwise analysis" to address these questions. For reasons discussed by Huberty (1989), stepwise analyses will not be considered here. Rather, a straight-forward multiple-analysis approach is favored. For example, for a situation involving 12 predictor variables, 12 eleven-variable analyses would be conducted, giving us a separate analysis "associated" with each of the 12 predictors. With each predictor deleted, a PDA is conducted using the remaining 11 predictors. With each predictor, then, is associated a hit rate generated by the other 11 predictors. A "small" hit rate (relative to the comparable hit rate based on all 12 predictors) would indicate that the associated variable is "important."

computer programs for PDA

Although there are at least four statistical packages (BMDP, SAS, SPSS, SYSTAT) available with potential for providing results related to a PDA, the packages of focus here are SAS (Version 6.1) and SPSS for Windows (Release 6.1.3). SAS has one procedure that is devoted to PDA, DISCRIM. The reason SAS DISCRIM was chosen is the vast array of PDA information yielded; SPSS DISCRIMINANT yields only one piece of information (about potential outliers) not obtainable from the other packages. Not all of the SAS DISCRIM output options will be considered herein. To learn more about DISCRIM, the reader may refer to a SAS manual or to four printouts (two of which are annotated) in Appendix B in Huberty (1994b). For completeness sake, reference to PDA procedures in the other three packages is briefly made as follows:

BMDP: 5M, 7M
SPSS: DISCRIMINANT
SYSTAT: MGLH

Printouts pertaining to PDA from the first two packages are given in Huberty (1994b).

An Illustration of a PDA

The illustration of a PDA used here is based on the research situation introduced earlier in this section. Recall that the problem dealt with predicting a decision of faculty members concerning departure from academia. The problem may also be viewed as one of *identification*; that is, the extent to which a faculty member may be identified with one group or another, where the three groups are defined as ("it" refers to "a permanent departure from academia"):

Group 1: I have given it serious consideration
Group 2: I have considered it, but not seriously
Group 3: I have not considered it

As mentioned earlier, the data were collected via a 1989 Carnegie Foundation survey. [After the study of this illustration, it may be informative to reread the earlier parts of this section on PDA.]

Sample Design

Following is a brief description of the sample design. A two-stage stratified random sample design was used to select college and university faculty. In the first stage, colleges (both four-year and two-year) and universities were selected; in the second stage, faculty were designated. Utilizing the Carnegie Foundation data bank of American colleges and universities, 306 schools were selected from a total of 2747 schools. Approximately 34 schools were selected from each of nine Carnegie classifications (from Research I universities to two-year schools). Within a classification, a school was selected with a likelihood proportionate to the size of its faculty compared to the other schools within that classification. The second stage of the sample design involved designating faculty at selected schools. Utilizing a data bank of American college and university faculty, maintained by the College Marketing Group, nearly 10,000 faculty members were designated for inclusion. The sample was divided equally among the nine Carnegie classifications.

Of the 9996 questionnaires which were mailed to college and university faculty across the country, 5450 faculty members returned their questionnaires, representing a return rate of about 55%. Across the nine classifications, the return rate ranged from 41% to 62%.

Data Preparation

Because 76 returned questionnaires were not usable, the data matrix initially considered had 5374 rows (i.e., 5374 faculty members in three groups of $n_1 = 1368$, $n_2 = 1661$, and $n_3 = 2345$) and 21 columns (21 predictor variable measures). However, because 15,652 (13.9%) of the total 5374 * 21 = 112,854 possible measures were missing, some "cleaning up" had to be done. To begin with, the columns in which there were 500 or more missing measures were deleted from further consideration. This resulted in a new p (i.e., number of predictors) of 12, reduced from 21. Second, the rows of the new 5374 * 12 matrix in which there were three[2] or more missing (out of 12) measures were deleted from further consideration; the number of such rows was 29. This resulted in a new matrix of 5345 rows (i.e., faculty) and 12 columns (i.e., predictors). But, this reduction still resulted in an incomplete data matrix; of the total 5345 * 12 = 64,140 possible measures, 623 (0.97%) were missing. If only a complete data matrix were to be used, then 544 faculty would need to be further deleted (because of some missing

[2]The numbers such as 500 and 3 are simply judgment calls to be made by the researcher. In this illustration there was a "luxury" of having a large number of analysis units (a maximum of 5374) relative to the number of predictors (a maximum of 21), with minimum-group-size-to-p ratios of 1368/21 and 1359/12.

predictor scores). The matrix involving 5345 faculty could, however, be completed by imputing the 623 missing values.

Prior to imputing missing values, it would be advisable to examine the available data for extreme or outlying faculty. This was done by visual inspection of the illustrative data set; no outlying data vectors were evident. There are some alternatives as to how data imputation might be accomplished, the simplest of which is to use the predictor mean for the respective group based on the available predictor scores. Other data imputation alternatives are briefly reviewed by Huberty and Julian (1994)—additional references are also given therein. For the current illustration, means of the available predictor scores for each respective group were used to replace the missing scores; a complete 5345 * 12 data matrix was thus constructed and utilized in subsequent analyses. With this data set, the group sizes were $n_1 = 1359$, $n_2 = 1654$, and $n_3 = 2332$.

In preparing the final data set for analysis, outlying score vectors must be taken into consideration. For the current example, an outlier is a faculty member whose vector of scores is "vastly deviant" from others. The meaning of "deviant" may vary from data set to data set and from researcher to researcher. One way to view deviancy is to use probabilistic distance. The SPSS DISCRIMINANT procedure calculates a "typicality probability" which reflects the distance of a faculty member's predictor vector of scores from the centroid of the group to which the faculty member is assigned. In this sense, probability refers to a tail-area of a chi-squared distribution, the probability of the observed score vector or one more extreme (Huberty, 1994b, pp. 79-80). The notion of typicality probability will be discussed later in this section. With the current data set, we used a tail-area cutoff of .020; that is, a faculty member's typicality probability of .020 or less indicates a potential outlier. This cutoff suggested 35 (2.6%) outliers in group 1, 32 (1.9%) in group 2, and 33 (1.4%) in group 3; a total of 100 (1.9%). What to do with the outliers is another issue. One thing that might be considered is whether or not a separate group exists that is comprised of some or all of the outlying units. This was not the case with this illustration. If it were the case, then a reexamination of group definitions needs to be done. One thing is suggested: Redo the analysis after deleting the identified outliers. Using some judgment, a decision may be made regarding outlier deletion. In the current situation, a reanalysis using 5245 rather than 5345 units showed no discernible differences in data condition assessments, descriptive information, or classification results (to be discussed later).

When analyzing survey data, as we are with the Carnegie data set, it is sometimes suggested that the observations be "weighted"—with some federal agencies, data weighting is required. Details for weighting strategies will not now be reviewed; suffice it to say that two purposes of weighting are to enhance accuracy and precision of statistical estimation when (1) disproportionate sampling is used in complex survey designs, and (2) nonresponses distort representativeness. Readable discussions of data weighting are given by Korn and Graubard (1991), Lee, Forthofer, and Lorimor (1989), and Lin and Schaeffer (1995). More technical discussions are given by Pfeffermann (1993) and Skinner, Holt, and Smith (1989). For the purposes of this illustration of PDA, it was decided *not* to use weighted data.

To carry out the PDA on this data set, some further data preparation was necessary. This preparation involved the scaling of 11 of the 12 predictors (see Table 1). Predictor 1 (gender) was scaled by using "0" for male and "1" for female. Predictors 2 and 4-10 have ordered categories; thus integer scaling was used for these predictors, as indicated in Table 1. The scaling of predictors 2, 3, and 11 was accomplished using the Fisher-Lancaster method mentioned earlier in this section[3]. Integer sealing was used with predictors 4-10. [Variables 13-21 in Table 1 were deleted from consideration because of missing data.]

Prior Probabilities

Another consideration that needs to be made before carrying out the PDA is the prior probabilities to use in the classification rule. The relative sizes of the three groups in the exemplary illustration were (before deleting *any* faculty members), respectively, about .25, .31, and .44. Because there was not complete confidence that the three group sizes reflected the corresponding population sizes, additional consideration had to be given to determining the priors to use. Two experienced researchers in higher education were consulted and asked to estimate relative sizes of the three corresponding populations. These two sets of proportions were "averaged" along with the above stated three proportions to arrive at the priors to use in this illustration: .15, .30, and .55.

Multivariate Normality

The form of the classification rule to be utilized depends upon two data conditions. One condition is the form of the joint distribution of the 12 predictors. If it is reasonable to assume that multivariate normality is approximated, then one could proceed with a normal-based rule. Normality assessment in the current case was accomplished by examining the normal probability-plots for the data in each group (obtained via the GRAPH option in SPSS for Windows). It was judged that the three plots did not indicate wild deviations from the straight-line plot for perfect multivariate normal distributions.

---

[3]The Fisher-Lancaster scale values were determined using all available category scores for each of predictors 2, 3, and 11 for the 5345 faculty. The missing scores on predictor 3, say, were replaced by respective group means of the scale values. It turned out that a comparison of the total-group 12 x 12 correlation matrices for the available-data (i.e., with missing data) matrix of $N = 5345$ faculty and for the imputed data matrix indicated that all corresponding coefficient values were within .01; the same result occurred when comparing the two 12 x 12 error correlation matrices.

TABLE 1. Predictor variables and how each is indicated and measured for the PDA example

Predictor Variable	*Indicators*	*Measures*
1. Gender	male, female	0, 1
2. Academic Rank	associate professor, full professor, no rank designated, other	F-L scale values
3. Kind of Appointment	tenure, no tenure, adjunct, visiting, other	F-L scale values
4. Impression of Student Suitability to Academic Life	Strongly Disagree...Strongly Agree	1, 2, 3, 4, 5
5. Salary Satisfaction	Excellent...Poor	1, 2, 3, 4
6. Satisfaction with Affirmative Action	Strongly Agree...Strongly Disagree	1, 2, 3, 4, 5
7. Dependency of Tenure on Publishing	Strongly Disagree...Strongly Agree	1, 2, 3, 4, 5
8. Perception of Importance on No. of Publications for Tenure	Very Unimportant...Very Important	1, 2, 3, 4, 5
9. Impression of Support of Administration on Academic Freedom	Strongly Disagree...Strongly Agree	1, 2, 3, 4, 5
10. Impression of Being Trapped in Terms of Advancement	Strongly Disagree...Strongly Agree	1, 2, 3, 4, 5
11. Ethnicity	Asian, Afro-American, Hispanic, Native American, Caucasian, Other	F-L scale values
12. Preparation for Teaching	time spent	No. hours per week
*13. Teaching Undergraduate Courses		
*14. Teaching Graduate Courses		
*15. Research and/or comparable scholarly activities		
*16. Scheduled Office Hours		
*17. Administrative Service		
*18. Consulting (with or without pay)		
*19. Service with curricular student activities		
*20. Supervising graduate teaching assistants		
*Deleted because of incomplete data		

Covariance Structure

Another data condition to consider before conducting the PDA is the comparability of the three group covariance matrices. A statistical test of equality of the three populations covariance matrices may be accomplished via the SAS or SPSS package. The F test yielded by SPSS DISCRIMINANT (as opposed to the SAS DISCRIM chi-squared test) is preferred by the current writers. For the data used in this illustration, $F(156, 55968464.8) \doteq 4.63$, $P \doteq .000$. As is obvious from the degrees-of-freedom values, the test is extremely powerful. That is, relatively small differences among the three sets of 12 variances and 66 covariances would be "detected." It had to be decided, then, whether to use a linear classification rule (which incorporates a pooled covariance matrix) or a quadratic rule (which incorporates three separate covariance matrices). A discussion of the linear-versus-quadratic decision will not be given here (see Huberty, 1994b, pp. 63-65, 259-260). For this illustration it was decided to use a *linear* rule. This decision is based on two factors: (1) even through a large F (of 4.627) transformation of the Box M criterion, the logarithms of the determinants of the three group covariance matrices (5.47, 4.17, and 3.62) were not judged to be radically different or that different from the logarithm of the determinant of the pooled covariance matrix (4.40); and (2) the precision (i.e., across-sample stability) of the hit rate estimates is known to be greater with a linear rule than with a quadratic rule. [Brief comparisons of the two sets of results will be made in a following subsection.]

Hit Rate Estimation

In a multiple regression context, the predictor variable composite is found in such a way so as to maximize the correlation between the composite scores and the criterion variable scores; that is, the sample R^2 is maximized. As such, then, R^2 is obviously a *positively* biased estimator for (i.e., it overestimates) its population counterpart. This is the situation because the data set used to determine the predictor composite is the same data set used to obtain the R^2 value. That is, an *internal* analysis is being, in effect, conducted. What may be used to reduce the bias is an *external* analysis—an analysis where the composite is determined using one data set and then is applied to another data set to obtain the desired estimate. [In regression analysis it turns out that a formula adjustment of the sample R^2 may be applied to obtain an unbiased estimator (see Huberty and Mourad, 1980).]

Now, a similar estimation problem exists in a PDA context. If a classification rule (say a set of predictor composites) is determined from a data set and then applied to that same data set, (positively) biased hit rates will result. Similar to the regression context, the classification functions are determined so as to maximize the group hit rates for the data on hand. So, then, an external analysis is needed to obtain unbiased hit rate estimates. [In the two-group PDA situation, hit-rate formula adjustments have been proposed—see Huberty (1994b, pp. 83-86).] A preferred external PDA is termed *a leave-one-out* (L-0-0) analysis. Suppose one has a total of 500 analysis units. One unit is "held back," and a rule is built on the remaining 499 units; and that rule is applied to the deleted unit. This process is repeated so that 500 rules are built and each rule is applied to the respective deleted unit; the 500 classifications are tabulated to obtain the group hit rate estimates. [See Huberty (1994b, pp. 88-90) for a more detailed discussion of the L-0-0 method of hit rate estimation.] L-0-0 results are obtainable via the SAS DISCRIM procedure.

For the purposes of this illustration, the misclassification costs across the three groups were taken to be equal.

The Analysis

In addition to the computer program for the Fisher-Lancaster scaling mentioned in an earlier section of this chapter, the computer programs used for this illustration are SAS DISCRIM (Version 6.1) and SPSS DISCRIMINANT (Windows, Release 6.1.3). To augment the latter, the DISCLASS macro may be used to obtain linear L-0-0 classification results.

Descriptives. The means, standard deviations, and error correlations for this illustration are given in Table 2. It may be noted that only one correlation was "high"—$r_{78} \doteq .74$, which is not too surprising because both of these predictors involve attitudes toward tenure. The remaining 60 correlations were "low," ranging (in absolute value) from .00 to .23 with a median value of approximately .04. The 12 sets of standard deviations are pretty much "in the same ballpark," with the greatest maximum-to-minimum ratio being $1.42/.91 \doteq 1.5$ for variable 10..

TABLE 2. Descriptive statistics for the PDA illustration (N = 5345)

Group Mean/ (s.d.)				*Error Correlations*										
Variable	1	2	3	2	3	4	5	6	7	8	9	10	11	12
1	1.35 (.48)	1.31 (.46)	1.25 (.43)	.00	.15	-.04	.07	.06	-.09	-.09	.01	.01	-.06	-.01
2	-1.30 (1.52)	-1.57 (1.36)	-1.71 (1.34)		-.23	-.05	.11	.00	-.02	-.04	-.01	.04	-.01	-.06
3	.67 (.82)	.69 (.74)	.61 (.70)			.01	.06	.04	.00	.03	-.04	.00	-.01	.04
4	3.71 (1.18)	3.52 (1.11)	3.31 (1.21)				.08	-.01	-.08	-.02	-.13	.14	.04	.02
5	3.12 (.88)	2.90 (.87)	2.60 (.90)					.09	-.02	.00	-.17	.23	.04	.06
6	3.02 (1.31)	2.77 (1.22)	2.61 (1.24)						.03	.04	-.22	.06	-.04	.02
7	3.73 (1.51)	3.74 (1.49)	3.71 (1.42)							.74	-.06	-.02	.07	-.07
8	3.63 (1.40)	3.60 (1.36)	3.53 (1.42)								-.11	.00	.06	-.06
9	3.37 (1.24)	3.69 (1.11)	3.96 (1.06)									-.16	-.05	-.03
10	2.86 (1.42)	2.20 (1.24)	1.47 (.91)										.02	.02
11	2.70 (1.23)	2.78 (1.34)	2.81 (1.30)											.04
12	10.92 (7.26)	10.78 (7.41)	10.36 (7.49)											

TABLE 3. Linear L-0-0 classification table for k=3

		Predicted Group 1	2	3	Total
	1	392 (28.8)	307	660	1359
Actual Group	2	196	293 (17.7)	1165	1654
	3	74	169	2089 (89.6)	2332
Total		662	769	3914	5345

Note. Group hit rates (in percents) are given in parentheses. Overall hit rate is 51.9 (2774/5345).

Group Results. The linear L-0-0 classification table is given in Table 3. The respective group hit rates are 28.8% (392/1359), 17.7% (293/1654), and 89.6% (2089/2332), while the total-group hit rate is 51.9% (2774/5345). It is obvious that with the 12 predictors considered, one cannot very well identify a faculty member with either of the first two groups. That is, with data on the 12 predictors, one would not be expected to predict very well if a faculty member has given serious consideration to permanently departing from academia (hit rate of 28.8%) or if a faculty member has considered departure but not seriously so (hit rate of 17.7%). For these two groups, we may want to ask if the hit rates are even as good as those hit rates that could be obtained by "chance." [Here, a proportional chance criterion is used.] A statistic that may be used to make this assessment is a standard normal statistic:

$$z = \frac{n_{gg} - e_g}{\sqrt{e_g(n_g - e_g)/n_g}},$$

where n_{gg} = the number of faculty assigned to group g, and $e_g = q_g n_g$ = the prior probability for group g times the size of group g. For group 1, n_{11} = 392, q_g = .15, e_1 = .15(1359) = 203.85, and $z \doteq 14.29$; the associated (one-sided) P value, which can be obtained from a standard normal distribution table, is less than .001. So, one might conclude that we have classified faculty into group 1 significantly better than would have been done by chance. The next "natural" question is: How much better than chance? To address this question, an "improvement-over-chance" index may be utilized:

$$I = \frac{H_o - H_e}{100 - H_e},$$

where H_o = the observed hit rate (in percent), and
H_e = the expected hit rate (in percent).

For group 1 we get

$$I = \frac{28.84 - 15.00}{100 - 15.00}$$
$$\doteq .163$$

Therefore, by using the linear rule developed, we can expect to make about 16.3% fewer classification errors than if classification was done by chance. It is judged that few (if any) researchers would consider this value of I to be "substantial." [See Huberty (1994b, pp. 103-108) for a more detailed discussion of the complete assessment procedure.] For group 2, $e_2 = 496.2$; because $n_{22} = 293 < 496.2$, the z statistic would not even be calculated. For group 3 (with a hit rate of 89.6%) we get $z \doteq 33.566$, $P << .001$, and $I = .768$. Therefore, by using the linear rule developed, we can expect to make about 76.8% fewer classification errors for group 3 than if classification was done by chance.

It may be noted that the lower bound of the (one-sided) 99% confidence interval for the number of hits in group 3 may be found using

$$n_{33} - z_{.99}\sqrt{e_3(n_3 - e_3)/n_3},$$

from which we get 2037.2. The lower bound of the (one-sided) 99% confidence interval for the group 3 hit rate would thus be approximately 87.4% (2037.2/2332).

As sort of a sidenote for the curious reader, it may be of interest to note that the quadratic L-0-0 group results were somewhat comparable to the linear L-0-0 results. The three quadratic hit rate estimates are, respectively, 29.7% (404/1359), 22.0% (364/1654), and 84.5% (1971/2332). The group 1 hit rate is about 0.9% higher than the linear hit rate, whereas group 2 hit rate is about 4.3% higher, and group 3 hit rate is about 5.1% lower. A statistical comparison of the linear results versus the quadratic results may be made using the McNemar statistic (Huberty, 1994b, pp. 108-110). The resulting value of the McNemar statistic is 2.191 with P=.139 (the referent probability distribution is a chi-squared distribution with 1 degree of freedom). Thus, it may be concluded that even in a statistical sense, the linear results are not "significantly" different from the quadratic results.

Faculty-by-Faculty Results. To study the classification results in a little more detail, the classification probabilities associated with individual faculty may be investigated. There are two types of faculty for which we might search. One type is an "outlier"—a faculty member assigned to a group but one who is very deviant from the typical member of the assigned group. A "typical member" is one whose vector of 12 scores is very close to the mean vector (i.e., centroid) of the assigned group. Such faculty could be identified by calculating a "typicality probability" for each faculty member assigned to each group (see Huberty, 1994b, pp. 46, 76-77). Even though "extreme" outliers may already have been deleted (see the subsection, Data Preparation), it may be informative to identify the type of faculty member whose group membership may be questionable.

The second type of faculty of interest is a fence-rider (or in-doubt case); one whose group assignment is not "clear-cut." A fence-rider is one whose assigned group posterior probability is close to that for another group. For this data set, two posterior probabilities were considered to be "close" if they were within .02 of each other. Of the 699 faculty from group 1 (see Table 3) who were assigned to group 1 or to group 2, there were only 55 fence-riders. Of the 489 faculty from group 2 who were assigned to group 1 or to group 2, there were only 54 fence-riders. Of the 2258 faculty from group 3 who were assigned to group 2 or to group 3, there were only 35 fence-riders. So, it is pretty clear that presence of fence-riders did not appreciably affect the separate-group hit rates. Characterizations of the 144 fence-rider profiles will not be attempted here; however, such characterizations may be substantive interest when classifying new faculty.

TABLE 4. Linear L-0-0 hit rates for 11-variable subsets

Variable Deleted	*Group 1 (Rank)*	*L-0-0 Hit Rate Group 2 (Rank)*	*Group 3 (Rank)*	*Total-Group (Rank)*
10	92 (1)	8.6 (1)	92.2 (12)	45.2 (1)
1	28.4 (7)	16.4 (7)	88.4 (1)	51.3 (7)
2	27.4 (7)	17.0 (7)	90.0 (6.5)	51.5 (7)
4	28.9 (7)	16.6 (7)	89.6 (6.5)	51.6 (7)
6	29.1 (7)	17.1 (7)	89.6 (6.5)	51.7 (7)
8	28.7 (7)	17.3 (7)	89.6 (6.5)	51.7 (7)
9	28.6 (7)	16.6 (7)	90.0 (6.5)	51.7 (7)
3	28.9 (7)	17.4 (7)	89.6 (6.5)	51.8 (7)
12	28.8 (7)	17.6 (7)	89.5 (6.5)	51.8 (7)
7	28.8 (7)	17.8 (7)	89.5 (6.5)	51.9 (7)
5	28.2 (7)	17.5 (7)	90.2 (6.5)	52.0 (7)
11	28.8 (7)	17.7 (7)	89.8 (6.5)	52.0 (7)
None	28.8	17.7	89.6	51.9

Relative Predictor Importance. As indicated earlier in this chapter, a primary purpose of conducting a PDA is to assess the predictive power of a set of predictors. If this is a purpose of the research example posed here, then it may very well be of interest to assess the relative predictive power of the individual predictors. To make this assessment, 12 eleven-predictor analyses (leave one variable out each time) were done. The results of these 12 analyses are reported in Table 4. It is clear from these results that predictor 10 (Impression of Being Trapped) is the most important predictor (because its deletion yields the lowest hit rate) when it comes to total-group predictive accuracy. For group 3, however, predictor 10 is the least important predictor. It should be recognized that in research situations where one predictor does not dominate all other predictors, the rank-ordering of the predictors may very well be more varying. As is perhaps obvious, the assignment of predictor ranks will be, to some extent at least, a judgment call. Different research situations may call for predictor rank-orders with respect to different group hit rates. Or, multiple sets of rank-orderings may be of substantive interest.

Predictor Selection. To start, it was decided to retain the following five predictors in this research situation: 2 (Academic Rank), 3 (Kind of Appointment), 5 (Salary Satisfaction), 9 (Impression of Support), and 10 (Impression of Being Trapped). It was reasoned that for this illustration, these five predictors would be judged to be "essential" by many researchers for predicting departure from academe[4]. The question of predictor selection, then, becomes one of which of the other seven predictors should be retained in developing the prediction rule. To address this question, all possible subsets of size 6, 7, ..., 11 were determined in terms of group 3 hit rates for each subset of each size. [Any other separate-group hit rate or the total-group hit rate could have been considered.] The best subsets of sizes 6, 7, ..., 11 are given in Table 5. It should be mentioned that the best subset of a given size was not much better than the second and third (and sometimes larger) best subsets. Further, for this data set, subsets larger than five predictors yielded hit rates no better than the selected (i.e., forced) subset of size 5. It may also be noted that the selected subset of size 5 yields a higher group 3 hit rate than does the complete set of 12 predictors—it is often the case that deletion of some predictors will yield a better prediction rule than that yielded by the complete collection of predictors.

TABLE 5. Group 3 hit rates for best predictor subsets

Additional Variables[a]	*Linear L-0-0 Hit Rates*
None	90.4
V7	90.4
V7, V8	90.2
V7, V8, V12	90.1
V6, V7, V8, V12	90.0
V1, V4, V6, V7, V12	89.9
V2, V4, V5, V7, V8, V12	89.8
V1, V4, V6, V7, V8, V11, V12	89.6

[a]Variables in addition to V2, V3, V5, V9, V10

Classifying New Faculty. It was mentioned in the earlier section (PDA Purposes) that there are two primary purposes and a secondary purpose of conducting a PDA. Analyses for the first primary purpose and for the secondary purpose have been discussed to this point. Now the second primary purpose of developing a prediction rule to use with new faculty will be discussed. The rule based on the five selected predictors, V2, V3, V5, V9, and V10, will now be used to classify some new faculty. The five-element predictor profiles of four "new" faculty to be utilized are as follows.

[4]It would be tempting for a researcher to use the variable-ordering results as in Table 4 to select an initial subset of predictors. Such an initial selection may be appropriate if the subset made "substantive sense" and, perhaps, was related to previous research. Whatever, a judgment call (with some rationale) would need to be made.

Faculty	*V2*	*V3*	*V5*	*V9*	*V10*
5346	1.40	0.41	3	3	5
5347	-0.66	1.14	3	4	4
5348	-2.41	1.29	1	5	1
5349	0.70	-1.18	2	4	4

Weights of the three linear classification functions (LCFs) were applied to the four profiles; the two largest LCF scores for the four new faculty are as follows.

5346	$LCF_1 = 14.31$	$LCF_2 = 13.48$
5347	$LCF_2 = 18.44$	$LCF_1 = 18.41$
5348	$LCF_3 = 12.53$	$LCF_2 = 11.09$
5349	$LCF_1 = 11.08$	$LCF_2 = 11.02$

The group assignments are indicated by the subscript on the larger LCF score. For example, faculty 5347 would be assigned to group 2 (those who consider departure from academia, but not seriously), but not decisively. It is more clear, perhaps, that faculty 5348 would be identified as one who would not consider departure (i.e., would be assigned to group 3).

If a researcher has access to the original set of faculty on whom the rule was based, there is a more straight-forward approach to classifying a new faculty member. With this approach, one simply includes the new faculty vectors of predictors in with the original set but with no group identification. The SAS DISCRIM procedure will calculate the three posterior probabilities of group membership for each new faculty, values of which may be used in making a group assignment (Huberty, 1994, pp. 112-113). For the four new faculty indicated above, the two largest (linear L-0-0) posterior probabilities (PPs) are:

5346	$PP_1 = .64$	$PP_2 = .28$
5347	$PP_2 = .40$	$PP_1 = .39$
5348	$PP_3 = .78$	$PP_2 = .19$
5349	$PP_1 = .36$	$PP_2 = .34$

The group assignments are indicated by the subscript on the larger PP value. These assignments are the same as those based on the LCF scores. Group assignments for faculty 5346 and 5348 would be fairly clear-cut, but not so for faculty 5347 and 5349. It is much easier to identify potential fence-riders, when it comes to group assignments, using posterior probabilities than using LCF score values.

In a practical, real-life, research situation, it would be desirable to update the classification rule when predictor scores on a sizable number of new faculty are obtained.

If it was decided to settle in on a five-predictor model, then it may be of interest to rank-order the five predictors (in terms of relative contribution to predictive accuracy). This may be simply accomplished by conducting five four-predictor analyses—linear L-0-0 analyses in the context of the current illustration. Such a reanalysis is necessary because any variable ordering is an ordering that is only relevant to the set of predictors being utilized in the final rule considered.

TABLE 6. Linear L-0-0 classification table for k=2

		PredictedGroup		
		1	*2*	*Total*
Actual Group	1	1597 (53.0)	1416	3013
	2	374	1958 (84.0)	2332
Total		1971	3374	5345

An Alternative Analysis. The results of the three-group analysis may suggest a little different research question to some: How well can we predict (using the 12 available predictors) whether a faculty member simply considers departure from academia or not? This would obviously call for a two-group analysis. Using the Carnegie data, we would have n1 = 1359 + 1654 = 3013 and n2 = 2332; respective priors used are .45 and .55. The group results for a linear L-0-0 analysis of these data are given in Table 6. The group 1 hit rate is 53.0% (1597/3013); this is more than twice the hit rate for group 1 and group 2 combined in the three-group analysis, which is (392 + 293) / (1359 + 1654) $\doteq$ 22.7%. The group 2 hit rate in the two-group analysis (1958/2332 $\doteq$ 84.0%) is about 5.6% lower than the hit rate (89.6%) for the same group in the three-group analysis. A comparison of the three-group results with the two-group results via the McNemar test yielded a chi-squared value (with 1 degree of freedom) of about 584.8 with $P << .0001$.

It may be concluded, then, that one could better identify faculty who simply would consider departing from academia period, rather than identify faculty who would either consider departure seriously or consider departure to a lesser degree. If consideration for departure is simply a yes-no situation, then it also looks like we could not as accurately identify those who say "no" as when the "yes" is qualified (as in the three-group situation). This conclusion is drawn because the hit rate for the "no" faculty (84.0%) in the yes-no situation is below the lower bound (87.4%) of the 99% confidence interval for hit rate of the "no" faculty when the "yes" is qualified. So, whether we consider a two-group prediction or a three-group prediction depends upon which type of faculty we are more interested in identifying.

If such a two-group design would be the one of interest, then the researcher would proceed as indicated in the following steps:

1. Consider predictor deletion; that is, delete those predictors judged not to contribute to high predictive accuracy;
2. Establish the final prediction rule;
3. Discuss group-prediction results and faculty-by-faculty results;
4. Rank-order the predictors; and
5. Consider the use of the final rule with new faculty.

Comments. The intent of the above analysis method was to illustrate some of the potentially interesting information that may be obtained and reported in a predictive discriminant analysis (PDA) research context. Illustrated also were some of the considerations that may be desirable to make in the analysis process. The how-to-do aspect of a PDA (i.e., specific use of computer programs) was not considered in detail (see, e.g., Huberty, 1994b, chaps. V-VIII). Also not considered to a great extent in the illustration are the substantive aspects (e.g., design, conclusions) of the research situation.

Summary

A summary of some considerations to be made in a study in which a predictive discriminant analysis (PDA) would be conducted is offered:

1. Representative sampling of units. Sampling method and sample description are essential.
2. Definitions of levels of grouping variable. Levels should be defined so that any potential analysis unit would clearly be associated with a single level.
3. Initial choice of predictor variables and their measures. Initial choice may be based on researcher knowledge and reasoning, and on previous research; predictor measures to be used should be clear, including any imposed scaling.
4. Search for outlying predictor vectors. Multiple searches may be made, some may be direct data eye-balling and some may be PDA-related.
5. Completeness of data matrix. A search for missing data should be made. Decisions regarding action subsequent to discovery of missing data are judgment calls and should be so indicated. Completion of the data matrix should be done in coordination with the search for outliers.
6. Assessments of multivariate normality and covariance matrix homogeneity. Graphical plots may be used for normality; n's, logarithms of covariance matrix determinants, and P value for Box M test may be reviewed.
7. Selection/deletion of predictor variables. Based on researcher judgment, some predictors may be retained "by default"; all-possible-subset analyses should be used.
8. Classification results, including outliers, fence-riders, hit rates, and chance classification. External classification, such as leave-one-out analyses, should be used. Examination of outlier and fence-rider score profiles may suggest comments regarding some special types of analysis units.
9. Predictor ordering. A set of all-but-one-predictor analyses should be conducted; a particular group hit rate may be the focus.

There is another analysis that may be of interest in a special circumstance. Suppose a prediction rule is developed on one sample of units. It may be of interest to compare the classification results for that sample against the results of applying the developed rule on another available sample. To make this comparison (for a particular group or for the total group), one would be comparing independent proportions—as opposed to comparing dependent proportions as was done above when comparing linear results against quadratic results in the same sample using the McNemar test. To compare the two independent proportions, one can use a chi- squared or standard normal statistic (Moore, 1995, p. 506).

3. DESCRIPTIVE DISCRIMINANT ANALYSIS IN HIGHER EDUCATION RESEARCH

Group Comparisons—An Introduction

Suppose one is studying the comparison of faculty in the four levels of academic rank with respect to the single response variable, perception of academic freedom. Suppose further that this variable may be measured on a continuum. A typical data analysis approach to this comparison is to use univariate analysis of variance (ANOVA)—this assumes that a legitimate academic rank comparison is one based on group perception means (and not on, for example, proportions). With this situation, academic rank plays the role of a grouping variable, also serving as an explanatory variable. On the other hand, the variable, "perception of academic freedom," plays the role of an outcome or criterion variable. It should be pointed out that even though (an omnibus) ANOVA was mentioned above as an associated analysis, more interesting analyses, perhaps, would address questions pertaining to the study of *contrasts* among the rank groups.

Let us complicate this situation some—and make it more realistic (?). Suppose the comparison among $k = 4$ academic ranks is to be made with respect to a collection of p outcome variables, and not merely a single outcome variable as in the case of an ANOVA. Now we have a multiple outcome variable situation. Analysiswise this might initially call for an omnibus multivariate analysis of variance (MANOVA). Just as in the ANOVA situation, it may now be more relevant to investigate multivariate *contrasts* among the $k = 4$ rank groups.

Let us first discuss the omnibus assessment of differences among the k mean vectors; that is, the omnibus comparison in a one-way layout will be considered first (by conducting a one-way MANOVA). [Contrast effects will be discussed later.] The basic assessment of testing the equality of the corresponding k population mean vectors is accomplished by using any one of four popular named test criteria—Wilks, Bartlett-Pillai, Hotelling-Lawley, and Roy (Huberty, 1994, pp. 183- 189). What is to be reported, then, is the value of the criterion of interest, a transformed test statistic value (for the first three criteria), a P value, and an effect size index value. If the P value is "small" and the effect size value is "large" (judgment calls are needed here), then it may be concluded that the k population mean vectors are different. Another conclusion that may be drawn is that there is a nonchance relationship between the grouping variable (as defined by the grouping variable levels) and a composite of the p outcome variables.

Just as with ANOVA, some data conditions technically need to be met to legitimately use any of the four test criteria mentioned above. Briefly, these pertain to score vector independence among the analysis units, multivariate normality of the outcome variable vector scores, and homogeneity of the k group covariance matrices. See Huberty (1994) for some detail on these conditions. It will be assumed outright that the outcome variable score vectors were independently obtained.

It will be assumed at the outset that a MANOVA yields group mean differences; that is, the MANOVA P value is small and the effect size index value is large. The discussion will center on how to "interpret" the resultant differences by using a collection of techniques that will be referred to as descriptive discriminant analysis (DDA). These techniques are also applicable in the context of a two-factor MANOVA where interest may be on factor interaction effects, main effects, simple effects, or contrast effects.

Exemplary DDA Research Questions

For the sake of introduction, examples of research questions will be restricted to situations involving only one-factor designs. Thus, exemplary questions will pertain to group differences or, equivalently, to grouping variable effects. Recent issues of journals related to higher education were reviewed to obtain the following research questions:

- How do college students grouped by "reflective judgment stage" differ with respect to a number of measures of critical thinking?
- How do various universities compare in terms of a number of measures of faculty attitudes toward teaching, research, administration, and work environment?
- In what way do ninth-grade students with decided postsecondary plans differ from those with undecided plans in terms of student background variables, characteristics, attitudes, ability to pay, et al.?
- With respect to various measures of research productivity and perceptions regarding research, how do universities in three levels of "publication performance" compare?
- In what ways are public university donors different from nondonors in terms of attitudinal variables, demographic variables, involvement variables, and philanthropic variables?

Although these questions, and many others found in the higher education literature, relate directly to a statistical group comparison (i.e., MANOVA), related interpretation concerns may be addressed by using DDA techniques.

DDA Purposes

As noted earlier, descriptive discriminant analysis techniques are useful for various interpretation purposes when studying the effects of some grouping variable(s) on a collection of outcome variables. These techniques would be useful when it is concluded that the grouping variable effects are generalizable. The exact composition of the set of DDA techniques is not definitive for all methodologists. Some techniques discussed in this chapter may not technically come under the DDA umbrella, but they will be included for the sake of "completeness."

The primary purpose for conducting a DDA is to identify outcome variable *constructs* that underlie the resultant effects of the grouping variable(s). The following form of a question is in the context of a single grouping variable: On *what* does the grouping variable have an effect? Or, with *what* is the resultant group separation associated? Suppose groups are found to have different mean vectors. These groups are different with respect to some construct(s) associated with the collection of outcome variables. Whether or not some meaningful construct(s) may be identified is, of course, highly dependent on the initial choice of the outcome variable collection.

There is another question that is often raised in the context of a multivariate analysis: What are the "important" outcome variables? In a MANOVA situation, this question may or may not be associated with DDA techniques. In this chapter, determining relative outcome variable importance will be considered as a secondary purpose of a DDA.

Computer Programs for DDA

Information pertaining to DDA obtainable via the four popular statistical packages varies considerably. The SAS package is the lone package with programs designed specifically for DDA; the programs are STEPDISC, CANDISC, and GLM. A program in the BMDP package—BMDP 7M—and the SYSTAT MGLH program yield some information for both a DDA and PDA. In the SPSS package, the DISCRIMINANT program yields information for both PDA and DDA, whereas the MANOVA program is for DDA only. Information obtainable from all four packages for both PDA and DDA is summarized in a table by Huberty (1994, pp. 21-22).

An Illustration of a DDA

Design and Data

For this illustration it was decided to again use the Carnegie survey data set. These data were collected from faculty members at nine types of institutions:

1. Research University I
2. Research University II
3. Doctoral Granting University I
4. Doctoral Granting University II
5. Comprehensive University/College I
6. Comprehensive University/College II
7. Liberal Arts College I
8. Liberal Arts College II
9. Two Year College/Institution

For our purposes, four groups of institutions were formed:

G_1: Research (types 1 and 2)
G_2: Doctoral Granting (types 3 and 4)
G_3: Comprehensive (types 5 and 6)
G_4: Liberal Arts (types 7 and 8)

A k = 4-level grouping variable was thus defined. The initial group sizes were n1 = 1267, n2 = 1315, n3 = 1212, and n4 = 1146, with N = 4940. The raw research question for this illustration was as follows: How do these four types of institutions compare? Of course, we need to specify the basis of comparison. This calls for a collection of outcome variables. To define the outcome variables, some 54 relevant survey items were selected. The intent was to pool items (in a substantively meaningful way) to arrive at item composites.[5] Before item composites were considered, it was necessary to clean up the 4940 x 54 data matrix.

The initial aspect of the clean-up pertained to outlying faculty available score-vectors. That is, are there some faculty whose scores were so extreme that they should be excluded from further consideration? Visual inspection did not reveal any such faculty.

Another aspect of the clean-up pertained to missing data. It was decided to drop any items for which there were 500 or more missing scores; there were five such items. Thus, 49 columns in the data matrix remained. Then the 4940 x 49 data matrix was examined for rows (i.e., faculty members) with seven or more missing item scores; there were 174 such faculty. We then had a 4766 x 49 data matrix. Also, there were three items that had a "Not Applicable" response option. There were 483 faculty who responded this way and thus were deleted. The raw data matrix considered, then, was one with N = 4283 (n_1 = 1090, n_2 = 1146, n_3 = 1072, and n4 = 975). [It should be noted that the sequence followed in deleting faculty and items may have an effect on the resultant data matrix to be used in the final analysis.] But there still remained some missing data. There were 1897 missing item scores which was less than 1% of the total number of possible item scores. The imputation method used was that of replacing the missing item score with the respective group mean. There was one item dealing with academic advising responsibility having unordered categories that had to be scaled prior to imputation. The Fisher-Lancaster method of scaling (mentioned earlier in this chapter) was employed. For this item, the respective group scale score mean was used to replace the missing scores. All of the other 48 items were scored using ordered categories varying from four to six categories. For these items, the miss-

[5]A second way of clustering items is to use some type of data reduction analysis. An analysis that is popular for this purpose is a principal component analysis. Details of a PCA are not given herein (see, e.g., Rencher, 1995, chap. 12). The use of such an analysis is not always clear-cut in terms of arriving at a reasonable number of interpretable components, which would then be considered the response variables for a PDA or a DDA. One potential issue is determining the number of components to retain. Another issue is labeling or naming each component. Still another issue is how to arrive at component (i.e., response variable) scores to use as input for a PDA or a DDA.

ing scores were replaced with the integer nearest the respective group mean. After data imputation was done, we had a complete 4283 x 49 data matrix.

Formation of item composites to represent outcome variables was completed on the basis of item "content" and professional judgment. We constructed seven such composites that were to be considered as seven outcome variables. Two items represented outcome variables by themselves. The final nine outcome variables are described in Table 7.

TABLE 7. Outcome variables and how each item within the composite is indicated and measured for the DDA example

Outcome Variable	*No. Items*	*Indicator*	*Measure*
1. Teaching Preparation Time	1	Time spent	No. hours per week
2. Institution Climate	7	Excellent...Poor	Sum of 1...5 for the seven items
3. Academic Advising Responsibility	1	Faculty, full-time advisors, student affairs professionals, others, no provisions	F-L* scale values
4. Institution's Entrance/ Graduation Standards	2	Much higher... much lower	Sum of 1...5 for the two items
5. Department's Entrance/Graduation Standards	1	Much higher... much lower	Score for item
6. Faculty Working Conditions	17	Strongly Agree... Strongly Disagree	Sum of 1...5 for the 17 items
7. Faculty Attitudes Toward Students	7	Strongly Agree... Strongly Disagree	Sum of 1...5 for the 7 items
8. Institution's Management Style	11	Strongly Agree... Strongly Disagree	Sum of 1...5 for the 11 items
9. Department's Management Style	2	Very Autocratic... Very Democratic	Sum of 1...4 for the 2 items
	49		

*F-L denotes Fisher-Lancaster

By examining the nine-element score vectors for the 4283 faculty, it was concluded that there were 89 potential outlying faculty members. The examination was done using a typicality probability cutoff of .020 (refer to the PDA illustration presented earlier in this chapter). Excluding the 89 faculty would reduce our N to 4194. The question then becomes, should the 89 faculty be included for the final analysis? There are at least two notions to consider in arriving at an answer to this question: (a) Would the 89 composite score vectors be those of "real" faculty members?, and (b) What effect on the analysis results does the inclusion of the 89 score vectors have? The score vectors of the 89 identified faculty were all judged to be those of "real" faculty. The analysis results to be compared with and without the 89 faculty pertain to data conditions, descriptive information, and DDA results. It should be realized that by mentioning comparison results, we could be accused of "getting ahead of the story" because some of these analyses are not specified until later. Results pertaining to group multivariate normality and group covariance matrix equality were virtually the same with and without the 89 faculty member outcome score vectors. The same conclusion was reached with respect to descriptive information (outcome variable means and standard deviations and error correlation matrices) and DDA results. Thus, there was no evidence that would suggest that the 89 faculty members should be deleted; the final data matrix used, then, is one with 4283 rows (for faculty) and 9 rows (for outcome variables).

Data Conditions

Two data conditions needed to be checked before proceeding with the data analysis. [Independence of the outcome variable score vectors was assumed.] One condition is the multivariate normality of the nine element score vectors in each of the four groups. Normality was assessed by examining the four normal probability plots. It was judged that deviancy from normality was minimal for all four groups. The other data condition that needed to be considered is homogeneity of the four group 9 x 9 covariance matrices. With the statistical tests readily available and with our group sizes of about 1000 each, this assessment is problematic. The test criterion considered is the Box M statistic that is built on the logarithms of the four-group pooled covariance matrix and the four group covariance matrices. For the current data, M = 1006.7 and the transformed F (135, 39181293) $\doteq$ 7.43, with $P \doteq$ .0000. It is obvious that these results are due to the "large" n values; the maximum difference of the matrix log determinants was only about 1.4. It was thus concluded that even though there appeared to be a *statistical* difference, we judge the difference not to be *substantial*. [Also, the univariate group variance differences did not appear to be anything near substantial, except, possibly for outcome variable 3.] Therefore, we proceeded with the data analysis under the assumption that the data conditions are satisfactorily met.

The Analysis

The computer software used to conduct the analyses for this illustration was SPSS for Windows (Release 6.1.3) and SAS (Version 6.1).

Descriptives. The means, standard deviations, and error correlations for this illustration are given in Table 8. The only three "high" error correlations are r_{18} = .71, r_{46} = .67, and r_{48} = .73.

Omnibus Test. For the current data, the Wilks lambda value was Λ = .859 with F (27, 12474.16) $\doteq$ 27.72, and $P \doteq$ .000. Before concluding that the group mean-vector differences are "real," it would be well to have some kind of effect-size index value. This value may be obtained by considering some linear composites of the nine outcome variables. The composites of interest are called *linear discriminant functions* (LDFs). With k = 4 groups and p = 9 outcome variables, it is possible to obtain three LDFs. The number of LDFs to retain for interpretation purposes may be determined in three ways (statistical tests, proportions of variance, and LDF plots; Huberty, 1994, pp. 211-216). It was decided to retain two

TABLE 8. Descriptive statistics for the DDA illustration (N = 4283)

Outcome	*Group Mean/(s.d.)*				*Error Correlations*							
Variable	*1*	*2*	*3*	*4*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
1	8.75	9.91	11.53	12.0	05	-.03	.51	-.02	-.02	-.01	.71	.01
	(5.9)	(7.2)	(7.6)	(7.9)								
2	15.08	14.77	14.11	11.91		.00	.34	-.26	.04	.00	.51	-.21
	(4.5)	(4.3)	(4.5)	(4.1)								
3	.96	0.93	1.10	1.17			.02	-.01	-.01		-.02	.03
	(0.7)	(0.7)	(0.5)	(0.3)								
4	115.05	116.34	117.81	116.11				-.01	.67	.39	.73	-.19
	(14.1)	(14.3)	(14.6)	(14.9)								
5	2.36	2.34	2.31	2.46					.01	.01	-.11	.10
	(0.7)	(0.7)	(0.7)	(0.6)								
6	38.12	38.25	38.58	38.39						.16	.11	-.30
	(8.2)	(7.8)	(8.2)	(8.3)								
7	18.72	18.88	18.94	18.80							.06	.02
	(3.8)	(3.7)	(3.7)	(3.7)								
8	72.4	73.7	74.9	72.5								-.08
	(10.2)	(10.9)	(10.9)	(11.3)								
9	5.1	5.1	5.1	5.5								
	(1.3)	(1.3)	(1.3)	(1.3)								

LDFs. With two LDFs, an effect size index is $\tau^2 = 1 - \Lambda^{1/2}$, which may be adjusted by taking into consideration the number of outcome variables and total sample size (Huberty, 1994, pp. 194-195). For the current illustration, the adjusted value is $\tau^2_{adj} = .067$. Thus, about 6.7% of the variability in the scores on the two LDFs may be attributed to institution type. Because this index has rarely, if ever, been reported in the applied higher education research literature (or any other research literature, for that matter), it is difficult to assess the magnitude of such a value. This assessment should be coupled with the assessment of the obtained P value to conclude that the groups are different; that is, we want a "low" P value and a "high" value of τ^2_{adj}. For the purposes of the current illustration, it was concluded that the effect of type of institution on the collection of nine outcome variables was "real."

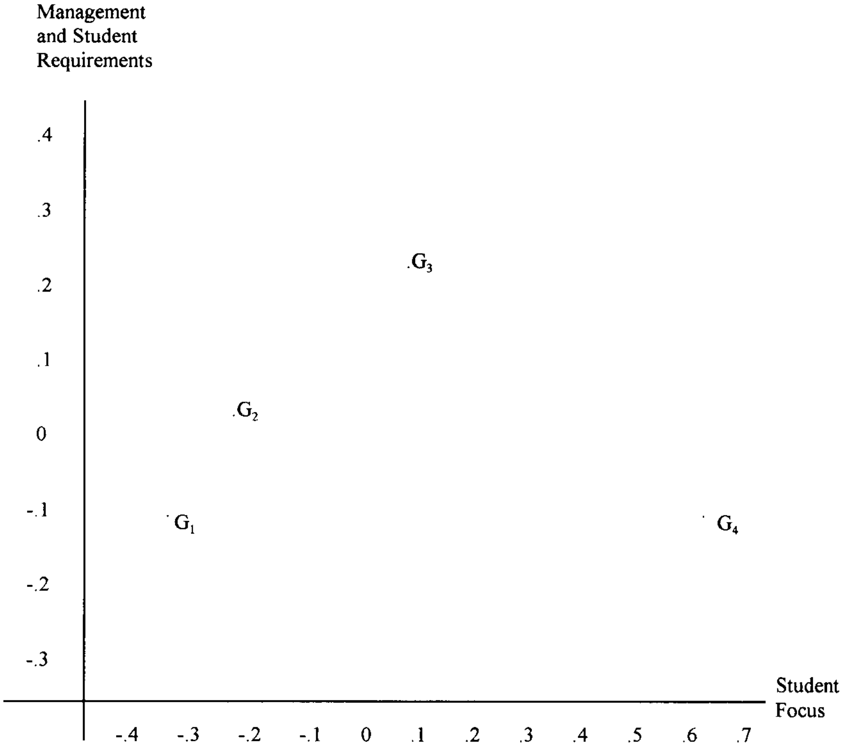
Underlying Omnibus Structure. If it is reasonable to conclude that the effect of institution type (as defined above) on the collection of nine outcome variables is "real," then the following question may be addressed: With *what* do we associate the institution-type effect? The "what" refers to what the two LDFs represent in a substantive sense. The weights for the LDFs are (mathematically) determined in such a way that intergroup differences are maximized with respect to the LDF scores; the LDF weights are "embedded" in the Wilks lambda criterion. The LDF weights determine, of course, scores on the LDF composites. If we have group separation with respect to these LDF scores, then the question of interest becomes the following: What do these LDFs represent in a substantive sense? These LDFs are conceptually similar to principal components. And, similarly so, to get a handle on what the LDFs represent, we can examine the nine variable-LDF error correlations, or the *structure* r's, for each LDF. The structure r's for the current analysis are given in Table 9. LDF_1 is dominated primarily by "Scope of Student Development" (r = -.73), and secondarily by "Teaching Preparation Time" (r = .44) and "Responsibility for Advisement" (r = .43). The label arrived at for the construct associated with LDF_1 is "Student Focus." Thus, it may be concluded that type of institution (as defined above) has an effect on "Student Focus." By examining LDF_2 structure r's, it may be concluded that type of institution also, but to a lesser extent, has an effect on "Management and Student Requirements," a combination of, basically, "Institutional Management Style" (r = .72), "Teaching Preparation Time" (r = .53), "Institutional Student Entrance/Graduation Requirements" (r = .49), and "Departmental Management Style" (r = .46).

Table 9. Omnibus structure *r*'s for DDA illustration

	LDF	
Outcome Variable	1	2
1. Teaching Preparation Time	**.44** (2.5)	**.53** (2)
2. Scope of Student Development	**-.73** (1)	.39 (6)
3. Responsibility for Advisement	**.43** (2.5)	.16 (8)
4. Institutional Student Entrance/ Graduation Requirements	.06 (7.5)	**.49** (3.5)
5. Department Student Entrance/ Graduation Requirements	.14 (5)	-.41 (5)
6. Faculty Working Conditions	.04 (7.5)	.13 (8)
7. Faculty Attitudes Toward Students	.04 (7.5)	.16 (8)
8. Institution's Management Style	.00 (7.5)	**.72** (1)
9. Department's Management Style	.33 (4)	**-.46** (3.5)

Note. The dominating values are boldfaced. Rank-orders are given in parentheses.

It may be of interest to see how the four institution types are related in a typological sense. To accomplish this, the LDF mean vectors (i.e., centroids) may be plotted in the space of the two retained LDFs. This plot is given in Figure 1. As is obvious from the plot, Liberal Arts institutions (G_4) are distant from the others with respect to LDF_1 which was defined as "Student Focus." With respect to LDF_2 ("Management and Student Requirements"), Liberal Arts, Doctoral Granting, and Research institutions are quite similar, with Comprehensive institutions (G_3) most distant from the other three.



Note. Institution types:
G_1: Research G_3: Comprehensive
G_2: Doctoral Granting G_4: Liberal Arts

FIGURE 1. Plot of institution centroids in the two-dimensional LDF space.

Omnibus Variable Ordering. In some research situations, it may be of interest to assess the relative importance of the outcome variables. The problem with this assessment is the interpretation of the word "important." Two interpretations pertain to: (a) LDF construct definition, and (b) group separation. With the first interpretation, there would be as many outcome variable orderings as there are LDFs. To arrive at the orderings, one simply uses the (absolute) values of the structure r's for each LDF. For the current illustration, the outcome variable rank-orders for the two LDFs are indicated in Table 9. The assignment of ranks involves judgment calls because of some "close" structure r values. Obviously, "Scope of Student Development" is most important in defining the construct "Student Focus"; "Teaching Preparation Time" ($|r|$ = .73) and "Responsibility for Advisement" ($|r|$ = .43) are next most important, with "Departmental Management Style" ($|r|$ = .33) and the remaining five variables contributing relatively little to the definition of "Student Focus." A variable importance sentence may also be written for the second construct, "Management and Student Requirements."

To arrive at an ordering in the sense of *overall* group separation (the second interpretation of "important"), a set of all-but-one-variable analyses is conducted. With nine outcome variables, there would be nine eight-variable analyses. That variable, when deleted, that yields the least separation (the highest Wilks lambda value) would be the most important variable. Results of the all-but-one-variable analyses are given in Table 10. It is clear that the least separation results when "Scope of Student Development" (Λ = .888) is deleted, with "Responsibility for Advisement" (Λ = .882) being a "close" second. Thus, these two variables are judged to be the most important in the overall separation of the four institution types. The other seven outcome variables are judged to be equally (un)important.

Table 10. Variable ordering results for the DDA omnibus illustration (N = 4283)

Variable Deleted	*Wilks Lambda*
Scope of Student Development	.888 (1.5)
Responsibility for Advisement	.882 (1.5)
Teaching Preparation Time	.867 (6)
Department's Management Style	.865 (6)
Institution's Management Style	.863 (6)
Faculty Attitudes Toward Students	.862 (6)
Institutional Student Entrance/ Graduation Requirements	.861 (6)
Faculty Working Conditions	.861 (6)
Departmental Student Entrance/ Graduation Requirements	.860 (6)
None	.859

Note. Proposed ranks are given in parentheses.

Contrast Analyses. In some—many?—group-comparison studies involving more than two groups, there is a natural interest in comparing only two of the groups, comparing one group versus an "average" of two other groups, comparing two averages of groups, or, in general, testing the null hypothesis that some linear composite of the k outcome variable mean vectors is zero. [This very same idea of group contrasts is applicable also in the single outcome variable (i.e., ANOVA) context.] A contrast is typically specified prior to conducting an omnibus MANOVA and DDA, but a contrast may also be specified after conducting the omnibus analysis; results of an omnibus analysis may, in fact, suggest some contrast effects to study. It is *NOT* necessary to delay contrast analyses until after a MANOVA is conducted. If contrast effects are of primary interest, then the researcher should feel comfortable in going directly to the investigation of contrasts. An omnibus MANOVA should be conducted *only* if the omnibus effects are of potential substantive interest (Huberty and Morris, 1989).

For the current illustration, we will investigate three contrast effects. The first contrast is based on the question: With respect to *what* are Research institutions different from Liberal Arts institutions? This would call for comparing G_1 and G_4; contrast weights are +1, 0, 0, -1. The analysis of this contrast (done via SPSS MANOVA) yielded a Wilks lambda value of .900, F (9,4271) $\doteq$ 53.50, $P \doteq .000$[6], and η^2_{adj} = .099 (η^2 is defined to be 1 - Λ; see Huberty, 1994, p. 195 for an adjustment formula). From this information one might conclude that Research institutions and Liberal Arts institutions are different, on the average, with respect to a *composite* of the nine outcome variables. This composite is, as discussed earlier, an LDF which is the "what" referred to in the associated research question stated above. To label this LDF, it's structure r's are examined (see Table 11). What accounts for the difference between Research and Liberal Arts institutions may be termed "Student Development" because of the dominant structure r for "Scope of Student Development" (r = .74).

Table 11. Structure r's for DDA contrasts

	Contrast		
Outcome Variable	G_1 vs G_4	G_3 vs $G_1 + G_2 + G_4$	G_2 vs $G_1 + G_3 + G_4$
1. Teaching Preparation Time	-.46	**-.60**	.30
2. Scope of Student Development	**.74**	**.62**	**-.64**
3. Responsibility for Advisement	.38	-.35	**.63**
4. Institutional Student Entrance/ Graduation Requirements	-.08	-.21	.00
5. Departmental Student Entrance/ Graduation Requirements	-.14	-.03	.12
6. Faculty Working Conditions	-.03	-.07	.04
7. Faculty Attitudes Toward Students	-.02	-.08	-.04
8. Institutional Management Style	-.02	-.23	-.11
9. Departmental Management Style	-.33	-.20	.30

Note. The dominating values are boldfaced.
Institution type:
G_1: Research
G_2: Doctoral Granting
G_3: Comprehensive
G_4: Liberal Arts

The second contrast question is: With respect to *what* are Comprehensive institutions different from a pooling of the other three types? This calls for a comparison of G_3 on the one hand and G_1, G_2, and G_4 on the other hand. Weights for this contrast would be +1, +1, -3, +1. The Wilks lambda value for this contrast is .944 and F (9, 4271) $\doteq$ 27.91, $P \doteq .000$[6], and $\eta^2_{adj} \doteq$ .050. We will proceed under the assumption that it is reasonable to conclude that the contrast effects are real. Based on the structure r's in Table 11, we came up with the label "Student Development and Teaching Preparation Time." This construct, then, is what underlies

[6]Because there were three contrast tests, the P value for each test was 3 times the observed tail area of each F distribution.

the separation of Comprehensive institutions from the other three types.

The third contrast question of interest is the following: What separates Doctoral Granting institutions from the other three types? Weights for this contrast are +1, -3, +1, +1. The Wilks lambda value is .971, and $F\,(9, 4271) \doteq 13.98$, $P \doteq .000^6$, and $\eta^2_{adj} \doteq .022$. Assuming "real" contrast effects, the structure r's for the lone LDF are examined (see Table 11). The construct underlying the separation reflected by this contrast may be labeled "Student Development and Advisement Source."

Implicit in contrast construct definition is an ordering of the outcome variables (as mentioned earlier). Another view of variable ordering involves a relative assessment of contrast effects. To order the outcome variables in this sense, one can conduct nine eight-variable analyses for each contrast. All-but-one-variable analysis results for the three contrasts specified above are given in Table 12.

TABLE 12. Wilks lambda values when variables are deleted for the three DDA contrasts

	Contrast		
Variable Deleted	G_1 vs G_4	G_3 vs $G_1+G_2+G_4$	G_2 vs $G_1+G_3+G_4$
1.Teaching Preparation Time	.901 (3)	.973 (6)	.985 (5)
2.Scope of Student Development	.922 (1)	.976 (2)	.984 (5)
3.Responsibility for Advisement	.909 (3)	.983 (1)	.987 (5)
4. Institutional Student Entrance/ Graduation Requirements	.895 (7)	.972 (6)	.985 (5
5. Departmental Student Entrance/ Graduation Requirements	.895 (7)	.972 (6)	.984 (5)
6.Faculty Working Conditions	.895 (7)	.971 (6)	.985 (5)
7.Faculty Attitudes Toward Students	.895 (7)	.972 (6)	.985 (5)
8.Institutional Management Style	.895 (7)	.972 (6)	.986 (5)
9.Departmental Management Style	.899 (3)	.972 (6)	.985 (5)
None	.895	.971	.983

Note. Proposed ranks are given in parentheses.

Now for a few comments regarding the constructs associated with the three contrasts. The development of college students involves more than one "aspect." Although its scope varies across institution types, some possible aspects are the following: cognitive, physical, moral, social, career, spiritual, personality, and educational. These aspects of student development emerge at all types of institutions in a variety of ways. Hence, it is not, perhaps, surprising that there is a relatively high correlation between "Scope of Student Development" and all three contrast LDFs (see Table 11). It might be concluded, then, that the characteristic of "Scope of Student Development" is one that would be prevalent in many types of institution contrasts. Not only does this characteristic play a major role in defining constructs that underlie the three contrast effects (as well as the omnibus effect), this characteristic is also considered relatively important from a separation standpoint. "Scope of Student Development," simply put, plays a major role in the separation among and between the types of institutions.

Variable Selection. In predictive discriminant analysis (PDA), predictor variable selection/deletion considerations are to be recommended in virtually every situation. In descriptive discriminant analysis (DDA), however, a rationale for outcome variable selection/deletion may be difficult to develop. The initial choice of a meaningful collection of outcome variables in a DDA will usually preclude the necessity of any subsequent deletion of variables. It can be readily discovered if a variable contributes little (relative to the other variables) to underlying construct definition or to group separation. If so, there is no apparent need to delete some outcome variables—unless, of course, one has an "extreme" situation of the smallest group size being, say, 50, and the initial number of outcome variables being 25. In such an extreme situation, one might conduct 25 ANOVAs and discard those variables yielding an F value of less than 1.0. High variable intercorrelations may also suggest some variable deletions. Reexamination from a common sense perspective of the outcome variable collection may suggest some deletions. The stepwise method of variable selection that is so prevalent in the applied literature is *not* recommended (Huberty, 1989; Huberty, 1994b, pp. 227-231).

Two-Factor MANOVA/DDA

All of the basic ideas of MANOVA and DDA in a one-factor context apply to the two-factor context. Linear discriminant functions (LDFs) are derivable for all effects of interest: interaction, main, simple main, and contrast. That is, construct definitions are possible for any effects of interest. Ordering of the outcome variables may be done for any type of effects. The *first* step in a two-factor analysis involves the investigation of interaction effects—unless there is a strong rationale stated for no interest in interaction effects, which would suggest a different-from-usual data analysis model. Next, other effects of interest may be investigated.

Summary

A summary of ideas that may be considered in conducting a MANOVA and a DDA is offered.

Pre-Analysis

1. Sampling. Explicit discussion of how analysis units are selected is expected of manuscript writers. If the units were randomly selected, then the collection of units from which the sample was selected should be described, and the author's interpretation of "random" should be stated explicitly. It is particularly helpful to provide some demographic information about the units so that the reader may assess sample representativeness. To state that the sample consisted of introductory psychology students is *not* very informative. In addition, adequacy of sample size should be assessed.

2. Group definition; i.e., definitions of the levels of the grouping variable(s). The definitions should be such that group membership of any potential analysis unit would be clear; that is, units should be classifiable into one and only one

group. If a multiple-factor design is used, a description of the way in which the cells are arranged (with cell size) should be included.

3. Outcome variables. Ideally the collection of outcome variables should constitute a variable *system* in the sense that the variables conceptually and substantively "hang together." This initial choice of variables may be based on substantive theory, previous research, expert advice, and professional judgment. The rationale used for including variables should be made clear. Explicit listing (e.g., in a table) of all outcome variables and how each is measured would enhance manuscript readability. Any use of data transformations should be reported. The reporting of the reliability of the measures for each outcome variable would be a real plus.

4. Outlying observation vectors. As is well known, a few outliers can "foul up" an analysis in surprising ways. An indication that a search for outliers was conducted and steps taken, if any, should be stated. For a discussion of outlier detection in psychology, see Orr, Sackett, and Dubois (1991).

5.Completeness of data matrix. The manner of handling missing data should be discussed. For example, see Roth (1994). [A second search for outliers should be conducted after the data matrix is completed.]

6.Data conditions. A brief discussion of the extent to which the available data satisfy the conditions of group multivariate normality and equal group covariance matrices should be given.

Analysis

7. Descriptives. There are three basic types of descriptive information for a k-group, p- variable MANOVA situation that should be reported: k means and standard deviations for each outcome variable, and the p x p error correlation matrix. One might also report a k x k matrix of Mahalanobis squared distance values. As a sidenote, another type of information that may be considered *descriptive* consists of the p univariate F values. This descriptive information may indicate to the reader some of the "strong" outcome variables, and, if an F value is less than 1.00, that variable would be contributing more "noise" than "signal." [Caution: *Univariate F* tests should not be used to assess relative variable contribution in a *multivariate* study.]

8. Statistical tests. For MANOVA main, interaction, or contrast effects, the following test information is suggested: criterion (e.g., Wilks) value, test statistic value (with degrees-of- freedom values), P value, and effect size value. If covariance matrices are not judged to be equal, then information for one or more Yao tests (Huberty, 1994, pp. 199, 203) should be given. Information for contrast effects tests would be the same as above for the omnibus effects tests.

9. Labeling of linear discriminant functions (LDFs). This information would be relevant if an argument is implicitly or explicitly made for approximate equality of group (or cell) covariance matrices. The number of LDFs to consider may be determined in one or more of three ways (statistical tests, propor-

tions of variance, and LDF plots). The retained LDFs may be interpreted/named/labeled by examining the LDF-variable correlations (sometimes called structure r's).

10. Optional information. Some optional information that may be reported includes LDF plots, outcome variable rank ordering, and outcome variable deletion—these details are reviewed by Huberty (1994b, chaps. XV, XVI).

Other Comments

11. References. There are at least four interrelated reasons for citing references: to convince the reader that the author is familiar with related literature; to indicate support for a stated argument or idea; to give the reader a source for more elaborate discussion pertaining to the stated idea; and to refer to previous related ideas. With regard to book references, two suggestions are advanced: use the most recent edition when applicable; and give page numbers unless referring to a general idea or philosophy. Following these two suggestions would provide an expected service to the reader, and would support the four reasons listed above. Another service that might be provided is to restrict the number of references—two or three should suffice—related to a particular argument or idea.

12. Computer software. Knowing which software package was used may be helpful for some readers, particularly beginning researchers. Not only should the statistical package be indicated (by year or version or release), but the particular programs or procedures that are used would be informative.

13. Methodologist as a collaborator. It may be helpful in some situations for the higher education researcher to have a quantitative methodologist or measurement specialist as a research collaborator. The methodologist would be someone who is currently *studying* the statistical method used; for example, someone who teaches the method to university-level classes or in professional workshops.

14. Writing hints. Of course, individual researchers have their own writing styles. But there have been some writings on writing that may be of help to some researchers (Aiken, 1994; Ehrenberg, 1982; Guion, 1983; Maxwell and Cole, 1995; Thompson, 1995). Although the second and fourth of these references were addressed to quantitative types, they contain some ideas from which higher education researchers could get helpful writing hints. The writer should place him/herself in the shoes of the reader as the writing is being done; simply put, we should write for the reader.

15. Finally, the research and data analysis methods used are viewed by some as a means to an end, and the *substantive* contribution is the most valuable aspect of a manuscript. True, the tail should not wag the dog! However, neither should this be a case of a dog with no tail or a very short one (and deformed, too?).

4. DISCRIMINANT ANALYSIS AS A SUPPLEMENT TO CLUSTER ANALYSIS

It is clear, hopefully, from what has been presented in this chapter to this point that discriminant analysis—PDA or DDA—techniques are applicable in multiple-group situations. That is, the data to be analyzed are based on multiple groups of students/faculty/institutions or whatever the analysis units are. As mentioned earlier, such groups must be well defined for discriminant analysis results to be interpretable. That is, any units that might be included in the original analysis and, related to a PDA, to which a prediction rule might be applied, must belong to one or another of the groups.

Cluster analysis (CA), on the other hand, involves techniques that are applicable to a single group of units. [See Bailey (1994), Blashfield and Aldenderfer (1988), Kaufman and Rousseeuw (1990), Lorr (1983), and Milligan and Cooper (1987) for some details on cluster analysis.] The main purpose of cluster analysis is to determine subsets of relatively homogeneous units, homogeneous with respect to a collection of response variable measures.

PDA and CA

Analysiswise, there may be a fruitful connection between discriminant analysis (DA) and cluster analysis (CA) in at least two different ways. First, let us consider how predictive discriminant analysis (PDA) techniques may be used to supplement a cluster analysis. Two main ideas come to mind here. One is to use classification results to aid in deciding the number of clusters to retain for interpretation and discussion. Using the cluster analysis response variables as predictor variables, one can build a classification rule with the data on hand and then use the rule to estimate hit rates for all of the clusters of a selected number, say, k. Then PDA hit rate estimates could also be determined for k-1 clusters and for k + 1 clusters (or for more numbers of clusters for that matter). The cluster typology for which the hit rates are most acceptable—a judgment call—would be a candidate typology to retain for interpretation purposes. Included in the PDAs, one might pay attention to the identification of possible outlying analysis units, and to possible in-doubt units. These two bits of information may be useful, along with the group hit rates, in determining the final cluster typology to retain. All of this will, of course, call for some judgment, common sense, and reasonableness.

The use of PDA in connection with cluster analysis may also be made as sort of an "external validation" analysis. Suppose a final cluster typology is decided upon using the original collection of response variables. Now consider one or more collections of additional response variables. Each extra collection may be used as predictor variables in assigning the original units to the clusters. This type of analysis was used by Harris and Kaine (1994), and Wastell and Gray (1987).

Suppose one conducts a cluster analysis and the resultant typology is one that "makes sense," appears to be generalizable, and is one in which the clusters are fairly well separated (i.e., in-doubt units are considered). Suppose further that the typology suggests a number of types of analysis units with which various "treatments" (e.g., varieties of counseling services) are associated. Then it may be of interest to identify cluster membership for a "new" unit so that this unit may be provided an appropriate "treatment." This identification may be accomplished by developing a classification rule via the use of PDA. For an example of developing a prediction rule for cluster assignment of school-aged children, see Huberty, DiStefano, and Kamphaus (1997).

DDA and CA

For a second connection between DDA and CA, let us consider how descriptive discriminant analysis (DDA) techniques may be used to supplement a cluster analysis. An issue with cluster analysis may be that of describing cluster separation with respect to the response variables. This may be viewed as a construct identification issue. As discussed in the DDA section of this chapter, construct definition is based on linear discriminant functions (LDFs). So, the decided-upon number of clusters will serve as levels of a grouping variable, and the response variables will serve as DDA outcome variables. The idea is to attempt to describe or define a construct structure of the response variables associated with the cluster typology. That is, given there are cluster differences, to *what*, in terms of the response variables, can the cluster differences be attributed? Stated another way, on what does the cluster typology have an effect? [The discussion in this subsection is based on the condition of equal covariance matrices.]

To answer such questions, one would start by determining the number of dimensions (i.e., number of LDFs) that is to be considered in the cluster typology interpretation process. Once this is done (using some judgment calls), to get a handle on what (in a substantive sense) the LDFs represent, the LDF-variable correlations—called structure r's—are examined. [This effort is similar to that involved in a factor analysis.] The larger structure r's indicate the response variables that are considered in labeling the construct. Such labeling would be built on researcher knowledge of the involved response variables and how the variables are measured and, of course, on some researcher judgment. There is another aspect of DDA that may help describe the typology that results from a cluster analysis. This aspect is a plot of the cluster centroids (i.e, mean vectors) in the space of the linear discriminant functions. Such a plot would depict the relative positions of the clusters—relative in the sense of general spatial location and in the sense of proximity. Huberty, DiStefano, and Kamphaus (1997) provide an applied illustration of the use of DDA techniques following a cluster analysis; see, also, Paykel and Rassaby (1978).

Finally, there may be some research interest in assessing the relative contribution of the response variables in distinguishing clusters. To make this assessment, a DDA-related approach to assessing variable ordering may be utilized. See

Fowlkes, Gnanadesikan, and Kettenring (1987) for an additional discussion of variable ordering in cluster analysis.

5. SOME ISSUES AND PROBLEMS

As in any area of data analysis, there are issues and problems in discriminant analysis that are, to some extent, unresolved. Some of these pertain to computer output, some to limitations of computer programs, some to analyst perceptions, and some to data analysis philosophy. Some issues in discriminant analysis discussed in this section—related to sample size, predictive discriminant analysis (PDA), descriptive discriminant analysis (DDA), response variable selection, and PDA versus DDA—are not resolved. Opinions and/or judgments are advanced for some, whereas for others, "resolutions" are merely in the form of suggestive questions.

Sample Size

The issue of desirable or minimum sample size pervades empirical research, in both univariate and multivariate data analysis contexts. For the former, researchers may refer to Cohen (1988), Desu and Raghavarao (1990), Kraemer and Thiemann (1987), and Odeh and Fox (1991) for guidance on desired sample size—by employing, of course, some professional judgments. For the multivariate context, specific guidance is much more limited; a good resource for some situations is Kres (1983). In the univariate statistical testing context, some specifications need to be made, such as probability of Type I and Type II error, and some type of effect size value. All three specifications are judgment calls. The same calls need to be made in multivariate contexts, some contexts of which pertain to statistical testing (e.g., canonical correlation analysis, MANOVA/DDA) and some of which do not pertain to statistical testing (e.g., cluster analysis, PDA). In any context, the number of response variables involved is an additional concern. Let us restrict the discussion here to the PDA and DDA contexts.

Predictive Discriminant Analysis (PDA)

It was written some years ago that to use multiple regression analysis one needs 100 analysis units; and a little later it was written where one needs 300 units. Of course, the idea of specifying a minimum sample size for any analysis is to "assure" that one obtains *precise* (i.e., limited variability over repeated sampling) estimates of parameters involved (assuming, of course, that estimator *bias* is not a concern). In multiple regression analysis, precision pertains to estimating the regression weights and the multiple correlation coefficient. But, precision is dependent on the number of predictors (p) as well as on the total sample size (N). Precision is also dependent on the magnitude of the true (i.e., the population) multiple correlation coefficient. That is, greater precision can be expected with a "large" N/p ratio along with a "large" true correlation coefficient.

The same goes for the PDA context. The minimum sample size needed for a "valid" PDA is dependent on the N/p ratio and the true hit rate of interest. Some *general* guidance has been given regarding the N/p ratio—some have said that the minimum group size should be $3p$ or larger (e.g., Huberty, 1994b, p. 96) making the total sample size at least $3kp$, where k is the number of levels of the grouping variable. In general, a larger sample is needed if a "small" hit rate (of interest) is expected than if a "large" hit rate is expected. [A "small" hit rate would be one that is better than chance by a "small" amount.] So, one might start by asking this question: What kind of a hit rate might I expect for my group definitions and my set of predictors? In general (greater specificity is mentioned below), if a "large" hit rate is expected, then one might start by using $3p$ as the minimum size for any group. If a "small" hit rate is expected, then one might start by using $5p$ as the minimum size for any group. If these minima are not reasonable from a data collection expense standpoint, then a reconsideration of the number of predictors (p) to be used may need to be made; that is, p may need to be decreased.

There are more specific considerations that may need to be made regarding minimum sample size for a PDA. These considerations pertain to the hit rate estimation method to be used—linear versus quadratic and internal versus external. If one is restricted to using the standard SPSS DISCRIMINANT program directly, a (linear) internal rule is employed which calls for a "large" sample. Here, the $5p$ criterion could be used with a "large" hit rate (of interest), and a $7p$ criterion might be used for a "small" hit rate. Under the assumption that the criterion specified applies to the smallest group size, the summary in Table 13 is offered. In the table it is assumed that a linear rule is used when covariance matrix equality is tenable; if a linear rule is used with clear covariance matrix inequality, then something like $7p$ and $9p$ might be considered for "large" and "small" hit rates, respectively. Assuming that an external rule is used (as it should always be), then one might be pretty "safe" if, in general, a $4p$ minimum group size criterion is used. It should be noted that the sample size suggested above are *not* all based on elaborate empirical evidence; rather, they are judgment calls.

TABLE 13. Smallest group size for a PDA.

		Rule Type			
		Linear Internal	Linear External	Quadratic Internal	Quadratic External
True Hit Rate	Large	$5p$	$3p$	$7p$	$3p$
	Small	$7p$	$5p$	$9p$	$5p$

Note. p = number of predictors

Descriptive Discriminant Analysis (DDA)

Some guidance for desirable sample size in MANOVA is given via tables in the aforementioned book by Kres (1983). To use these tables, five bits of information must be specified: α, β, p, k, and effect size. For example, with $\alpha = .05$, $\beta = .30$, $p = 8$, $k = 4$, and a moderate effect size, an n (number of analysis units *per group*) of 70 is called for—a total of $N = 280$ units. With this example, the *common group size* is approximately $9p$. For $\alpha = .01$, $\beta = .30$, $p = 8$, $k = 3$, and a large effect size, $n \doteq 50$ is advised; that is, $N \doteq 150$. With this example, $n \doteq 6p$. Scanning across the Kres tables for MANOVA (developed by J. Läuter), one might conclude that the smallest group size should range from about $6p$ to about $10p$, less for very large effect sizes, large α values, large β values, and few groups, and more for opposite conditions.

If one relies on the Kres tables, the sample sizes called for in the MANOVA context are somewhat larger than those suggested in the PDA context. Of course, those two contexts are different in terms of required specifications; values of α and β do not pertain to a PDA context, and the two "effect size" concepts are different.

There are no known sample size guidelines that are specific to DDA. It is suspected that, if pressed, most researchers would use the MANOVA sample size guidelines for the DDA context. There *may* be a difference in the guidelines for MANOVA and for DDA. Basic interest for the latter is in linear discriminant functions (LDFs). The concern of sample size, then, pertains to estimation of eigenvalues and of eigenvector elements. The study of sample-size determination for accurate and precise estimates of these parameters is virtually nonexistent. A related question is the following: How big a sample is needed to correctly determine the number of LDFs to retain for interpretation of group separation? Even though little guidance on desirable sample size is available explicitly for DDA, it seems reasonable to follow the general recommendation given above for MANOVA: smallest n ranges from $6p$ to $10p$.

Predictive Discriminant Analysis

Rule for New Units

Typically, the group-membership prediction rule that is considered for use with new units is the rule determined from the data on hand. Typically, too, the form of this rule is that of a set of k (the number of groups) classification functions—either linear or quadratic predictor variable composites, including the appropriate prior probabilities. This is what is suggested in the writings of statisticians and methodologists. There is an issue that is not known to have been considered when suggesting a data-based rule. This issue pertains to in-doubt analysis units or "fence- riders." If in-doubt units are to be ignored in hit rate estimation, should they also be ignored in determining the rule to be used with new units? That is, suppose m analysis units (out of the total of N) are identified as in-doubt units. Then, should the rule to use with new units be that based on N-m units or on N units? One thing that might be tried is to develop two rules, one on N units and one on N-m units, and try to make sense out of the comparative results by using some judgment and common sense. One thing that might be considered is whether or not a large percent of the m fence-riders assigned to a particular group emanate from one given group, or, whether or not a large percent of the m fence-riders are assigned to one particular group. Of course, other possibilities may arise.

There is another question to be addressed when classifying a new unit (using whatever rule that makes sense). This question pertains to the group assignment and possible characterization of the new unit. That is, is the new unit clearly a member of a particular group? Or, to what extent does the new unit "look" like the typical member of the group to which it is assigned? Is the new unit a fence-rider? An outlier?

As stated earlier, the typical rule form for new units is that of a set of classification functions. But with this form, it would be quite difficult to identify potential in-doubt units and outlying units. If, however, one has access to the original data set, complete classification information is obtainable through the use of SAS and SPSS. By merely annexing the original data with the observation vectors of the new units, one can simply use the LIST option in the SAS PROC DISCRIM statement—of course, there would be no group identification for the new units. There is a second method of classifying new units via SAS. First, use DATA ORIGINAL to create the data set used in developing the classification rule. Second, PROC DISCRIM is run with the option OUTSTAT = CUBS, say. Third, use DATA CHECK to create a data set consisting of the observation vectors for the new units. Fourth, run PROC DISCRIM with the options DATA = CUBS and TEST DATA = CHECK. [CUBS is the name given to the data set.]

With both of the two methods of classifying new units, either a linear or quadratic rule may be used, and either a resubstitution or leave-one-out analysis (using the SAS CROSSVALIDATE option) may be conducted. Any of these analyses will yield the k posterior probabilities for each new unit—thus, potential in-doubt units may be identified. Direct identification of outliers via SAS is not possible, however.

Whether or not a new unit is an outlier may be determined using the SPSS package. With this package we are limited to a linear internal (i.e., resubstitution) analysis via the DISCRIMINANT procedure, or a linear L-0-0 analysis via the DISCLASS macro.

Hit Rate Estimation

The common approach to the estimation of group hit rates is to use something termed a *counting estimator*. With this approach, for each analysis unit an index value for each group is determined—either a classification function score or a posterior probability value. Then to obtain the hit rate for, say, group 1, one simply counts the number of units for which the group 1 index value is maximum, and divides this number by the number of units in group 1. This is how the entries in classification tables yielded by BMDP 5M and 7M, by SPSS DISCRIMI-

NANT, and the default tables of SAS DISCRIM are obtained. If these counting estimates are based on an external analysis such as a leave-one-out analysis, then the hit rate estimates are quite *accurate*. That is, there generally is little resultant basis—over repeated sampling one can expect that the long run mean hit rate estimate is close to the true hit rate. But, some question has arisen regarding the *precision* of the counting estimator. That is, the stability of the counting estimator over repeated sampling has been questioned (see, e.g., Hora and Wilcox, 1982). Thus, what has been proposed is an alternative to the counting estimator that may be termed a maximum-posterior-probability (M-P-P) estimator. With the M-P-P estimator, rather than counting and obtaining a proportion of hits for, say, group 2, one obtains a mean of the posterior probabilities of all units assigned to group 2. A discussion of the M-P-P estimator, with examples, is given by Huberty (1994b, pp. 90-93). The M-P-P estimates are available via the SAS DISCRIM program with the POSTERR option. In fact, the SAS DISCRIM program will yield leave-one-out (L-0-0) results for the M-P-P method by using the CROSS-VALIDATE option. It turns out that one can also obtain something called "stratified" M-P-P estimates.

So, what hit rate estimation method should be used? Although the M-P-P estimator appears to be a little less accurate (i.e., more biased than the L-0-0 estimator), this deficiency (if there really is one in general) may be overcome by using a M-P-P/L-0-0 estimator. This estimator combination uses the L-0-0 method to estimate the posterior probabilities and uses the M-P-P method to obtain the group hit rate estimates. So, then, the choice seems to be the L-0-0 counting estimator versus the M-P-P/L-0-0 estimator. Of course, for very "large" samples (i.e., "large" minimum group size), either the counting L-0-0 estimator or the M-P-P/L-0-0 estimator should well serve the researcher. But then, what does "large" mean? Earlier it was mentioned that a minimum sample size, in general, is one for which the smallest group size is greater than four times the number of predictors. A reasonable conclusion at this point in time is to simply use the L-0-0 counting estimator, provided the minimum sample size is used.

Nonnormal Classification Methods

For data analysis in general, data distribution assumptions are the crux of parametric estimation methods. In PDA, the estimates of interest are the posterior probability values and the values of the weights in the (linear or quadratic) classification functions. To obtain these values (given in the "standard" PDA computer outputs), a particular data distribution form is assumed; that of multivariate normality. There are three approaches one can take with regard to data distribution form while conducting a "standard" PDA. One approach is to ignore the technical distribution requirements and "plod on" with the analysis. There are no statistical tests in a PDA that require multivariate normality for valid interpretation. However, if the multivariate normality condition is not met, then the posterior probability estimates and the discriminant function weight estimates will be less than optimal. The big concern here is, perhaps, less with the lack of estimate optimality than with the consistency of estimates across repeated sampling. If the researcher is comfortable with the representativeness of the sample of units in use, then it may be reasonable to assume that similar classification results would be expected with repeated sampling, however lacking in optimality.

But, there may be a nonnormal rule that will yield higher hit rates for the sample on hand, and can be expected to yield comparable hit rates over repeated samples. There are a fair number of nonnormal alternatives (see Huberty, 1994b, chap. X); only two will be mentioned here. This discussion is limited to variables that are measured using a continuum, as opposed to variables measured using categories. Suppose it is not reasonable to conclude that the data vectors are anywhere close to following a multivariate normal distribution, and that one is not willing to "plod on" while ignoring the technical requirement. One alternative, then, is to transform the data. A transformation that has exhibited some promise in a PDA context is the rank transformation. The use of "fancier" data transformations (e.g., square root, logarithm, arcsine) has not been shown to be too fruitful. The second alternative to consider that does not depend on any theoretical distribution form is a "nonparametric" one. A nonparametric analysis that is obtainable via the SAS package is something called a *nearest-neighbor analysis*. The jury is still out with respect to the confidence one can place on nearest-neighbor PDA results.

So, what does one do with data vectors that clearly are not in the multivariate normal arena? First of all, assuming one has a representative sample, the advice given here is to have a *large* sample—where the total sample size, N, is greater than, say, $7kp$—and proceed with the standard package normal-based analysis.

Outliers and In-Doubt Units

It should go without saying that every data set should be examined for anomalies prior to any statistical analysis. At the outset of an analysis, the anomalies to search for include missing data and outlying observation vectors. The missing data issue will not be discussed in this chapter—see, for example, Huberty (1994b, pp. 267-271) and Huberty and Julian (1995).

Outliers. With regard to outliers, one can check for miscoded data and for potential univariate outliers; that is, each column in the data matrix (corresponding to the individual predictor variables) should be checked. [Assuming the predictor scores were properly coded, one might check to see if there are, for interval-scaled scores, any that are, say, three standard deviations or more away from the mean. Such scores can be "flagged" as possible (univariate) outliers.] Identification of multivariate outliers is a bit more involved—see, for example, Rencher (1995, pp. 116-120). A point should be made at the start of this discussion: A univariate outlier need not be a multivariate outlier, and a multivariate outlier need not be an outlier with respect to any single predictor. So, how might potential multivariate outliers be detected in a PDA context? A suggestion follows. Retain all predictor observation vectors that are "legitimate," that is, those except for miscoded scores. Suppose an analysis unit is assigned (using whatever classification rule that is judged appropriate) to group 2. An outlier, then, may be defined as a unit that is far from the centroid of group 2. Although, "far" may be determined in terms of distance, it may be more easily determined in terms of probability. There is a question to be addressed: Given the unit is assigned to group 2, how likely is it that the unit belongs to group 2? This question may be addressed in terms of the extremeness of the unit's observation vector relative to the group 2 centroid. This extremeness is assessed using what may be termed a "typicality probability." This probability, as determined by the SPSS package, is simply a tail area of a chi-squared probability distribution; that is, it is the probability of getting the unit's observation vector or a vector more extreme. These probabilities are denoted as P(D/G)—"D" stands for "data"—by SPSS DISCRIMINANT.

Definition-wise, this may be acceptable. But, how are outliers specified? This is another one of those judgment calls. If P(D/G) = .002, then, clearly, we have an outlier. How about if P(D/G) = .05 or .10? This one part of a PDA calls for some thought and some snooping. Suppose for a given group there are a fair number of potential outliers, all with P(D/G) values less than, say, .06. Typologically, are they all in the same "area"? If a "substantial" number of units are in close proximity, maybe there is really one additional group/population. Are the score vectors for all potential outliers legitimate? Have many of them been assigned to the same group? Might some of the (extreme?) outliers be deleted from the data set for the final analysis? If, in a preliminary analysis phase, covariance matrix heterogeneity was concluded, might that result be due to the presence of outliers?

There are no simple answers to the above stated questions. Most, if not all, call for judgment calls. The issue of outlying units is of interest when it comes to interpreting PDA results, both for initial results and for assigning new units.

In-Doubt Units. Suppose an analysis unit is assigned to group 2 with an estimated posterior probability of .862. Remembering that the sum of the k probability estimates is unity, most researchers would agree that the group assignment is fairly clear-cut. And if a unit in a replication sample had a similar predictor vector, then that unit would be confidently assigned to group 2. But now suppose that a unit had a posterior probability estimate of .394 for group 1 and .387 for group 3. This unit would be assigned to group 1, but not with overwhelming confidence. If the same predictor vector for a unit resulted in repeated sampling, then that unit may very well be assigned to group 3. That is, confidence in group assignment would be lacking. Such a unit may be identified as an *in-doubt unit*, or a *fence-rider*. Of course, the labeling of an in-doubt unit is a judgment call. One can use .03 as a "margin of error," or .02, or .01, etc. One would expect that the greater the group separation (less overlap), the fewer the number of in-doubt units.

It is possible to conduct a PDA so that a minimum posterior probability estimate must occur for group assignment to be made. This may be accomplished via SAS DISCRIM with the THRESHOLD command. Suppose THRESHOLD = .45 is used; then for a unit to be assigned to any group, the associated posterior probability of membership in any group must be at least .45. Such an analysis may be done as in initial analysis, or as a follow-up to an analysis with which no THRESHOLD was used. An important choice may need to be made. If an intent of the research is to develop a rule to use with new units, which rule is used (see the subsection, Rule for New Units)—that based on all N of the original units, or that based on N-m where m is the number of in-doubt units? Assuming that the inclusion of the in-doubt units in the sample(s) contributes to the representativeness of the sample(s), then the rule to be used with new units should be based on data that included the predictor vectors of scores for the in-doubt units.

How might the identification of in-doubt units enhance the interpretation and discussion of PDA results? The presence of such units may distort the hit rate estimation process. For example, a large portion of in-doubt units may be assigned to a given group, thus inflating the hit rate estimate for that group. All of those assignments may be correct (who knows?); on the other hand they all may be incorrect (who knows?). So, one may want to report two hit rates (or, error rates), those based on data that include the predictor scores for the in-doubt units, and those based on data that exclude in-doubt analysis unit score vectors. The latter error rates may be termed "serious error rates." Such error rates may be compared across groups for the study on hand, or across studies.

Descriptive Discriminant Analysis

Standardized LDF Weights

It is common knowledge that the magnitude of an X-variable weight in a (linear) multiple correlation analysis is dependent on the metric used in measuring the X variables. To "make the weights comparable," at least some methodologists suggest using weights based on standardized X-scores. The comparability of the weights applicable to standardized scores may be questioned on the grounds that comparisons should take into consideration the estimated standard errors of the weights as well as the weights themselves (see, e.g., Montgomery and Peck, 1992, p. 158).

In DDA there are linear discriminant function (LDF) weights for the Y variables, and so-called standardized LDF weights may also be obtained. Some comments will now be offered regarding the use of LDF weights in "interpreting" results of a DDA. To start, the LDF weights are only determined up to a constant of proportionality. That is, if -0.973 and 2.349 are weights obtained for Y_1 and Y_2 using one computer program, then 1.946 and -4.698 may be obtained for the respective weights using a different computer program. [For that reason alone, numerical signs of the LDF weights are of limited interpretative utility.] Also, unlike the multiple correlation situation, the probability distributions of the LDF weights are not known. Therefore, direct estimation of LDF weight standard errors is not possible—

although a resampling estimation approach, such as the bootstrap, is a possibility. The reason that estimated standard errors are mentioned is that for one "interpretation" purpose, these estimates are judged to be essential.

What does it mean when one attempts to "interpret" DDA results? It appears that there are two interpretation issues. One is to name or label each LDF. An LDF is a linear composite of the Y variables and, as such, may represent a meaningful construct that underlies the resultant effect (as assessed via a MANOVA). Two approaches have been advanced for naming an LDF. One is to use something called *structure r's*. The structure r for variable Y_i and LDF_j is a (Pearson) correlation between Y_i and LDF_j; that is, r_{ij} denotes the correlation between the ith outcome variable and the jth LDF which is a linear composite involving the ith outcome variable. Thus, r_{ij}^2 reflects the amount of variance shared between Y_i and LDF_j. It might be argued that if r_{ij}^2 is "high," then whatever the Y_i measures represent and whatever the LDF_j measures represents have *something* in common. Typically there are additional Y variables with "high" structure r's. So then, there may be, say, three or four "somethings" that when coalesced by the substantive researcher, a construct emerges. [This is what a researcher does who employs factor analysis.] A second approach to naming an LDF has been advanced by some writers (e.g., Harris, 1993; Rencher, 1995, p. 315). This approach is based on the so-called standardized LDF weights (see, also, Rencher, 1992). The naming of a construct represented by an LDF is herewith based on Y variables that have "large" standardized weights (ignoring signs). This approach is *not* the one recommended by the current writers.

The second interpretation issue pertains to relative Y-variable contribution or ordering of the Y variables. The approach commonly used in multiple regression analysis to order the variables is based on the magnitude of the so-called standardized regression weights. This same approach is very often used by applied researchers in a DDA context. And it is the approach advocated by some statisticians/methodologists (e.g., Rencher, 1995, p. 315; Hair, Anderson, Tatham, and Black, 1995, p. 206). It is *not* the approach advocated by the current writers—see the next subsection.

In sum, the current writers would maintain that standardized LDF weights are of very limited interpretive value (Huberty, 1994b, pp. 231-234). Other writers, however, retain the option of using these weights for the two purposes mentioned above, that of naming LDFs and that of outcome variable ordering (see, e.g., Rencher, 1995, pp. 315-318; Stevens, 1996, pp. 264-265).

Outcome Variable Ordering

It is maintained that the big issue that remains when it comes to determining the relative variable importance is the meaning of "important." That is, with respect to *what* is one variable more important than another? To espouse the use of standardized LDF weights in a DDA context, one *might* be thinking about the relative contribution to the LDF score. Of course, then, one would present as many variable orderings as there are LDFs that are retained for interpretation purposes. But is that meaning of "important" relevant for interpretation purposes? Perhaps it is. It might be argued that the LDFs are determined so that group separation is maximized with respect to LDF scores and, therefore, focusing on relative contribution to LDF score values may be reasonable.

Another view of relative variable contribution pertains to naming the LDFs. It was mentioned above that the preferred (by some) naming approach is based on structure r's. So then, if "importance" refers to naming LDFs, a reasonable variable ordering index would be (the absolute value of) the structure r. If, say, the magnitudes of Y_2 and Y_7 structure r's for LDF_2 were largest, then it might be concluded that Y_2 and Y_7 were the most important for the naming of LDF_2. Of course, one would report as many outcome variable orderings as there are LDFs used in the results discussion.

There is yet a third view of the variable ordering problem. With this view, an outcome variable is considered relatively important if it contributes greatly to the effect of interest. "Effect" here may pertain to an omnibus grouping variable effect or to a contrast effect. Consider, for example, a three-group design and the effect of interest is the mean difference of group 2 and group 3. An outcome variable, Y_i, would contribute to the difference if the two-group separation decreased considerably when Y_i was deleted. So, then, for a p-variable problem, p analyses would need to be done, each involving p-1 variables. That variable, when deleted, with which the *smallest* separation is associated would be considered the *most* important variable. This set of all-but-one-variable analyses is easily done via the statistical packages (see Huberty, 1994b, pp. 231-232). Even though the analyses are easily accomplished, the actual variable ordering might very well call for some researcher judgment—this is so because two or three variables may have index values that are numerically "close" (see Huberty and Wisenbaker, 1992). An attractive aspect of this all-but-one-variable approach is that the associated conception of variable importance is applicable to many, if not all, multivariable analyses, including PDA (Huberty, 1994b, pp. 126-130). One question, in general, is addressed: How well can I do without a variable? The "do" varies from analysis to analysis.

An interesting approach to the outcome variable ordering problem has been advanced by Thomas (1992). What is proposed is an index termed a "discriminant ratio coefficient" (DRC) that is, in one form, simply a product of the corresponding structure r and standardized LDF weight. It may be noted that DRCs are also proposed to be used for construct naming, and for outcome variable selection, the topic to be addressed shortly.

Unequal Covariance Matrices

Nearly all the discussion of discriminant analysis in this section involves *linear* discriminant functions. That is, it was assumed that the group covariance matrices were approximately equal. Such equality is what leads to the linear composites (i.e., LDFs) that are named or labeled for substantive interpretation. Now consider this question: How is the inter-group structure determined when it is clear that the condition of equal covariance matrices is not satisfied? Herein lies the problem. If the homogeneity condition is not satisfied, then the usual MANOVA criteria (e.g., the Wilks lambda) are not legitimate; similarly for the derivation of LDFs. With unequal covariance matrices, group mean differences may be assessed in a contrast context by using the Yao test (Huberty, 1994b, pp. 199-200, 202-203). But there is no known way of defining inter-group structure when group differences are tested under the heterogeneity condition.

Response Variable Selection

The process of selecting/deleting variables—the how-to-do—in either a PDA context or DDA context will not be delved into here in any great detail. Detailed coverage is given by Huberty (1994b) for both PDA (pp. 118-126) and DDA (pp. 227-231). Rather, what is discussed here is the issue of whether or not some variables should be deleted in developing a "model" of prediction (PDA) or of group separation (DDA). In PDA, the decision of predictor selection does not appear to be an issue. Just do it! Invariably across prediction situations, a better classification rule can be developed with fewer than the original number of predictors—"better" in the sense of higher hit rates.

Variable selection in a DDA situation, however, is an issue. Scanning the applied literature in many fields of study will find that the researchers very often attempt to delete some outcome variables. Now if it may be assumed that the researcher(s) spent time and thought in choosing the original collection of variables, then it is not clear why there should be any interest in deleting some variables. If the original variable choice resulted in a meaningful *system* of variables, then that is the collection to study. Remember, a major purpose of a DDA is to determine a variable *structure* that underlies the group separation. If some outcome variables do not contribute to construct definition, that is informative in itself, and may be determined by examining the structure r's. Why redefine the variable system? Sure, there will be times when the original system may be too "large." But if all variables are initially judged to be relevant to the related substantive theory, then they should all be retained for analysis and interpretation purposes. An exception to this may be dealt with in a very simple manner. An initial analysis step may involve multiple univariate ANOVAs. If any ANOVA F values are less than 1.0, then those variables might be excluded from further study on the basis that all they are contributing is "noise."

PDA Versus DDA

Because "discriminant analysis" and "discriminant function analysis" are viewed as rather nondescriptive expressions, the more specific expressions of "predictive discriminant analysis" (PDA) and "descriptive discriminant analysis" (DDA) are preferred. It is judged that these latter expressions specify the purpose and type of analysis that is or was to be done. It is recognized that some computer output information pertaining to PDA and DDA is available for *both* a PDA and a DDA—by using, for example, BMDP 7M and SPSS DISCRIMINANT. But that does not justify a "mix-up" of the two types of information. The various aspects that differentiate PDA from DDA are summarized in Table 14. All nine bases of comparison have been implicitly or explicitly referred to in this chapter. It may be pointed out that a PDA is applicable to data based on a one-way layout, whereas DDA results may very well be utilized with a design that involves more than one grouping variable.

TABLE 14. PDA versus DDA

	PDA	*DDA*
1. Research concern	Prediction of group membership	Description of group separation
2. Variable roles:		
Predictor(s)	Response variables	Grouping variable
Criterion (ia)	Grouping variable	Response variables
3. Response variable composite	LCF	LDF
4. Number of composites	k	min(p, k - 1)
5. Preliminary analysis concerns:		
Equality of covariance matrices	Yes	Yes
MANOVA	No	Yes
6. Analysis aspects of typical interest:		
Variable construct(s)	No	Yes(!)
Response variable selection	Yes(!)	Maybe
Response variable ordering	Yes	Yes
7. Criterion for variable selection/ ordering	Classification accuracy	Group separation
8. Research purpose	Practical/theoretical	Theoretical
9. Interest in generalizability	Yes	Yes

Context: k groups of units, p response variables.

There may be some empirical research situations in which one might want to employ *both* PDA techniques and DDA techniques. To do so, however, it must be assumed that during the data collection phase, determiners of group definition are conceptually independent of measures on the response variables. First, consider the use of DDA-related results as a supplement to PDA results. Here, the researcher may simply want to *describe* group differences; statistically assessing group differences (with respect to the response variables) via MANOVA would not be called for. Second, consider the use of PDA results as a supplement to DDA results. Having concluded that there are group differences (via MANOVA), it may make sense to consider an overall (or separate-group) hit rate as an effect-size index value. The research situation should dictate whether only a DDA be conducted or only a PDA, or both.

In writing up the results of a multiple-group analysis, there would usually be an emphasis on the use of DDA *or* PDA. And if one set of results is used to supplement the other, an emphasis on the supplemental nature of the results should be made explicitly.

6. COMMENTS

The potential for the use of discriminant analysis—predictive (PDA) and descriptive (DDA)—appears to be considerable in higher education research. There are research issues dealing with group-membership prediction (or, identification) of college students and faculty where the groups of interest are naturally-existing groups. Information available via a PDA is potentially abundant: identification of outliers, identification of in-doubt analysis units, hit rates, hit rate assessments, predictor selection, predictor ordering, plus more specific information. PDA methods may also be utilized in conjunction with cluster analysis.

DDA information is also quite useful: linear discriminant functions (LDFs), LDF plots, outcome variable ordering, et al. Whereas PDA deals basically with practical problems, DDA applies more toward substantive theory issues, at least in a descriptive sense. It is the study of grouping variable effects to which DDA techniques apply. DDA techniques are applicable to research situations in which group membership is manipulated as well as in situations involving naturally-existing groups.

The intent of this chapter was to review and apply techniques of both PDA and DDA for higher education researchers. Specific details regarding formulation, computer programs, and computer printouts have not been a focus (see Huberty, 1994b).

References

Aiken, L.R. (1994). Some observations and recommendations concerning research-methodology in the behavioral sciences. *Educational and Psychological Measurement* 54: 848-860.

Bailey, K.D. (1994). *Typologies and Taxonomies.* Beverly Hills, CA: Sage.

Blashfield, R.K., and Aldenderfer, M.S. (1988). The methods and problems of cluster analysis. In J.R. Nesselroade and R.B. Cattell (eds.), *Handbook of Multivariate Experimental Psychology.* New York: Plenum.

Cliff, N. (1987). *Analyzing Multivariate Data.* San Diego, CA: Harcourt Brace Jovanovich.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New York: Academic Press.

Ehrenberg, A.S.C. (1982). Writing technical papers or reports. *The American Statistician* 36:326-329.

Desu, M. ., and Raghavarao, D. (1990). *Sample Size Methodology.* Boston: Academic Press.

Fowlkes, E.B., Gnanadesikan, R., and Kettenring, J.R. (1987). Variable selection in clustering and other contexts. In C.L. Mallows (ed.), *Design, Data, and Analysis.* New York: Wiley.

Guion, R.M. (1983). Editorial comments from the new editor. *Journal of Applied Psychology* 68: 547-551.

Hair, J.F., Anderson, R.E., Tatham, R.L., and Black, W.C. (1995). *Multivariate Data Analysis.* Englewood Cliffs, NJ: Prentice Hall.

Hand, D.J. (1986). Recent advances in error rate estimation. *Pattern Recognition Letters* 4:335-346.

Harris, G., and Kaine, G. (1994). The determinants of research performance: A study of Australian university economists. *Higher Education* 27: 191-201.

Harris, R.J. (1993). Multivariate analysis of variance. In L.K. Edwards (ed.), *Applied Analysis of Variance in Behavioral Science*. New York: Dekker.

Hora, S.C., and Wilcox, J.B. (1982). Estimation of error rates in several population dis-

criminant analysis. *Journal of Marketing Research* 19: 57-61.
Huberty, C.J (1989). Problems with stepwise methods: Better alternatives. In B. Thompson (ed.), *Advances in Social Science Methodology* (vol. 1). Greenwich, CT: JAI Press.
Huberty, C.J (1994a). Why multi-variable analyses? *Education and Psychological Measurement* 54: 620-627.
Huberty, C.J (1994b). *Applied Discriminant Analysis*. New York: Wiley.
Huberty, C.J, and Barton, R. M. (1989). An introduction to discriminant analysis. *Measurement and Evaluation in Counseling and Development* 22: 158-168.
Huberty, C.J, DiStefano, C., and Kamphaus, R.W. (1997). Behavioral clustering of school children. *Multivariate Behavioral Research* 32: 105-134.
Huberty, C.J, and Julian, M. W. (1995). An ad hoc analysis strategy with missing data. *Journal of Experimental Education* 63: 333-342.
Huberty, C.J, and Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin* 105: 302-308.
Huberty, C.J, and Mourad, S. A. (1980). Estimation in multiple correlation/prediction. *Educational and Psychological Measurement* 40: 101-112.
Huberty, C.J, and Wisenbaker, J. M. (1992). Variable importance in multivariate group comparisons. *Journal of Educational Statistics* 17: 75-91.
Kaufman, L., and Rousseeuw, P. J. (1990). *Finding Groups in Data*. New York: Wiley.
Korn, E.L., and Graubard, B.I. (1991). Epidemiologic studies utilizing surveys: Accounting for the sample design. *American Journal of Public Health* 81: 1166-1173.
Kraemer, H.C., and Thiemann, S. (1987). *How Many Subjects? Statistical Power Analysis in Research*. Newbury Park, CA: Sage.
Kres, H. (1983). *Statistical Tables for Multivariate Analysis: A Handbook with References to Applications*. New York: Springer-Verlag.
Lee, E.S., Forthofer, R.N., and Lorimor, R.J. (1989). *Analyzing Complex Survey Data*. Newbury Park, CA: Sage.
Lin, I-F., and Schaeffer, N.C. (1995). Using survey participants to estimate the impact of nonparticipation. *Public Opinion Quarterly* 59: 236-258.
Lorr, M. (1983). *Cluster Analysis for Social Sciences*. San Francisco: Jossey-Bass.
Maxwell, S.E., and Cole, D.A. (1995). Tips for writing (and reading) methodological articles. *Psychological Bulletin* 118: 193-198.
McLachlan, G.J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*. New York: Wiley.
Milligan, G.W., and Cooper, M.C. (1987). Methodological review: Clustering methods. *Applied Psychological Measurement* 11: 329-354.
Montgomery, D.C., and Peck, E.A. (1992). *Introduction to Linear Regression Analysis*. New York: Wiley.
Moore, D.S. (1995). *The Basic Practice of Statistics*. New York: Freeman.
Odeh, R.E., and Fox, M. (1991). *Sample Size Choice*. New York: Dekker.
Orr, J.M., Sackett, P.R., and Dubois, C.L.Z. (1991). Outlier detection and treatment in I/O psychology: A survey of research beliefs and an empirical illustration. *Personal Psychology* 44: 473-486.
Paykel, E.S., and Rassaby, E. (1978). Classification of suicide attempts by cluster analysis. *British Journal of Psychiatry* 133: 45-52.
Pfefferman, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review* 61: 317-337.
Rencher, A. C. (1992). Interpretation of canonical discriminant functions, canonical variates, and principal components. *The American Statistician* 46: 217-225.
Rencher, A.C. (1995). *Methods of multivariate analysis*. New York: Wiley.
Roth, P.L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology* 47: 537-560.

Skinner, C.J, Holt, D., and Smith, T.M. F. (1989). *Analysis of complex surveys*. New York: Wiley.

Stevens, J. (1996). *Applied Multivariate Statistics for the Social Sciences*. Hillsdale, NJ: Erlbaum.

Thomas, D.R. (1992). Interpreting discriminant functions: A data analytic approach. *Multivariate Behavioral Research* 27: 335-362.

Thompson, B. (1995). Publishing your research results: Some suggestions and counsel. *Journal of Counseling and Development* 73: 342-345.

Wastell, D.G., and Gray, R. (1987). The numerical approach to classification: A medical application to develop a typology for facial pain. *Statistics in Medicine* 6: 137-146.

Faculty Demography: Exploring the Effects of Seniority Distributions in Universities*

James C. Hearn
University of Georgia

and

Melissa S. Anderson
University of Minnesota

Most faculty and administrators are aware of the potential effects of organizational seniority distributions only through personal experience or anecdote. A professor knows which colleagues joined the department before his or her own arrival and which came later, and most faculty have a general sense of their colleagues' ages. They may talk about the special problems a department faces when its faculty are nearly all close to retirement or nearly all recently hired. Few have considered, however, the many ways seniority distributions might affect their institutions. Likewise, as higher-education analysts seek insight into structures, processes, and interpersonal relationships in colleges and universities, they often overlook seniority distributions and their interactions with other organizational factors. Because many departments, institutions, and systems are currently experiencing notable changes in their seniority distributions, better understanding of the implications of those changes seems especially valuable now.

Faculty seniority distributions are an aspect of the organizational demography of higher education. The large literature on demographic issues in organizations considers several factors related to seniority, including age, years of service, and level of position in the hierarchy (e.g., see Pfeffer, 1983). For higher-education faculty, we conceptualize seniority distributions as encompassing age distributions

*An earlier version of this review was presented at the Annual Forum of the European Association for Institutional Research, Zurich, Switzerland, August, 1995. The authors gratefully acknowledge generous grant support from the Spencer Foundation for the research. Marvin Peterson made especially helpful suggestions after reading an earlier version of the manuscript. James Eck provided diligent research assistance.

J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*

within academic units as well as distributions of such organizationally specific factors as years of service, rank, and tenure status. Linking these factors is their attachment to individuals rather than units and their relatively easy aggregation and measurability. Also linking them, at least among the more senior faculty, is their relative imperviousness to policy action: each of the various factors we label demographic is, for senior faculty, largely outside the scope of organizational intervention at the individual level. Junior rank in higher education can certainly be quite tenuous, but a tenured full professor is extraordinarily difficult to demote or dismiss. Among the veteran faculty of our colleges and universities, rank and tenure status tend to be every bit as predictable from year to year as age, race/ethnicity, or gender.

The first part of this chapter explores the nature of seniority in the distinctive context of academe. The next section of the chapter examines three arenas of potential organizational effects of seniority distributions in that context. Organizational analysts have long studied the effects of age, rank, and experience distributions within a variety of settings, and we draw on those earlier investigations of organizational demography in this section. We then suggest some further challenges for demographic analysis of academic institutions. We conclude with a discussion of the importance of demographically oriented research for theory and practice.

OVERVIEW

A significant demographic challenge looms for higher education. In the 1950s and 1960s, the ranks of faculty grew swiftly, due to extraordinary increases in enrollment and funding. In the 1970s and 1980s, growth slowed and faculty hiring greatly decreased. As a consequence, the age, rank, and experience distributions of academic faculty became increasingly tilted to the upper range (Clark and Lewis, 1985; Lozier and Dooris, 1987; Breneman and Youn, 1988; Hacker, 1992). Table 1 traces this progression in the age distributions of faculty. Between 1975 and 1992, the ratio of faculty aged fifty or older to those under 40 tripled. Table 2 shows a similar pattern in rank distributions: the only rank to grow proportionately between the 1970s and the early 1990s was full professor.

Now, in the mid-1990s, many of the faculty hired during the boom years are nearing retirement. In North America and elsewhere, these retirements arrive at a time of potentially increased demand: the number of eighteen-year-old youth is once again rising (ending the "baby bust" years), and growth in the number of older students continues. The pace of faculty retirements is expected to accelerate well into the early 21st century, and institutions increasingly need to hire new faculty to replace them. As a consequence, faculty age, rank, and experience distributions are beginning to tilt to the lower range, and many institutions and academic departments are moving from a "top-heavy" condition to a "bottom-heavy" condition.

TABLE 1: Faculty Age Distributions for U.S. Higher Education, Selected Years

Faculty Age	*1968*	*1972*	*1975*	*1979*	*1983*	*1987*	*1992*
% 30 or less	13	7					
% 31-35	17	15					
% 36-40	17	16					
% 41-50	28	31					
% 51-60	16	21					
% Over 60	8	8					
% 29 or less			11	2	1	1	1
% 30-34			14	13	9	7	7
% 35-39			18	23	17	15	13
% 40-44			14	18	22	20	17
% 45-49			13	15	16	21	19
% 50-59			19	22	24	26	31
% 60 or older			10	8	11	12	13
Ratio of 50 or older to under 40	.5(est.)	.8(est.)	.7	.8	1.3	1.7	2.1

Note: Data are for full-time instructional faculty. Columns do not always total to 100 because of rounding or, in the 1972 data, non-response. There is no single source for data on faculty age distributions over time, but the data here from varied sources are acceptably comparable. Source for data for 1968 and 1975 is Stadtman (1980). Source for data for 1972 is Bayer (1973). Source for data for 1983 and 1987 is Hammond and Morgan (1991). Source for data for 1992 is U.S. Department of Education (1995).

TABLE 2: Faculty Rank Distributions for U.S. Higher Education, Selected Years

Faculty Rank	*1964*	*1972*	*1978*	*1980*	*1987*	*1992*
% Instructor, Lecturer, etc.	21	24	25	24	20	23
% Assistant Professor	31	25	26	25	23	23
% Associate Professor	23	24	24	25	24	24
% Professor	25	26	25	26	33	31
Ratio of Professor to Assistant Professor	.8	1.0	1.0	1.0	1.4	1.3

Note: Data are for full-time instructional faculty. Columns do not always total to 100 because of rounding. The "Instructor, Lecturer, etc." category includes those holding no academic rank. Source for data for 1964 is U.S. Department of Health, Education, and Welfare (1965). Source for data for 1972 is Bayer (1973). Data for 1978 are derived from data in U.S. Department of Education (1981). Data for 1980 are derived from data in U.S. Department of Education (1982). Source for data for 1987 is U.S. Department of Education (1993). Source for data for 1992 is U.S. Department of Education (1995).

Of course, there will be exceptions to the pattern. Cost considerations are forcing many institutions and systems into delaying replacements for retiring professors, or hiring part-time rather than full-time, tenure-track faculty.

Some observers have therefore come to doubt the magnitude of the overall influx of new junior faculty. Unquestionably, there are substantial disciplinary, institutional, and regional differences in the pace and extent to which the tilt will take place. Nevertheless, most analysts of the problem and most administrators concur that, overall, striking change is coming in the demographic profile of U.S. academic departments, in the direction of faculty who are younger, lower in rank, and less senior professionally. A 1994 American Council on Education survey (cited in *The Chronicle of Higher Education*, 1994) counters any vision of a shrinking full-time faculty with no slots for assistant professors: the survey suggested that 49 percent of senior higher-education administrators in the U.S. reported a net gain in 1993-94 in their institution's number of full-time faculty positions and another 34 percent reported no significant change in that number, yet 53 percent of the same group reported either stability or loss in their institution's total number of tenured faculty members. Clearly, the ranks of faculty in the U.S. are already changing and will change even more dramatically in coming years. This emerging demographic transition is of increasing theoretical and practical importance to higher-education systems.[1]

In dealing with this transition, college and university leaders may obtain guidance from general organization theory and from research focused on other organizational settings. Haveman (1995, p. 586) has observed that "The demography of organizations—the distribution of the members of employing organizations along such dimensions as length of service, age, and race—is a flourishing area of investigation for organizational scholars." In a much-cited 1983 review essay, Jeffrey Pfeffer identified demography as a critical and often overlooked organizational factor, evaluated alternative conceptualizations of demography as an organizational variable, and provided a conceptual framework relating demography to a wide range of other organizational factors. Subsequently, organizational demography has attracted extensive attention among other researchers, and many additional empirical studies have been produced. As a whole, the growing body of research on the topic suggests strongly that a unit's demographic profile is a significant, influential *organizational* variable (see Stewman, 1988; Bacharach and Bamberger, 1992; Mittman, 1992). That is, demographic factors have notable organization-level consequences and are not merely the aggregated demographic characteristics of the individuals within the organization. As Bacharach and Tolbert (1992, p. ix) have noted, "a growing number of researchers have recognized that the context of social relationships, defined by the distribution of status characteristics among organizational members, often determines the impact of the characteristics on individual and organizational outcomes."

[1]The trend is not unique to the U.S. Barbara Binder, the head of Swiss science policy, has noted (1995) that one-third of the senior faculty in Switzerland is expected to retire in the decade between 1995 and 2005.

Little of this work has penetrated higher-education research, however, and the interaction between demography and such distinctive organizational aspects of academic organizations as professional autonomy and the preparation and socialization of new faculty (see March and Olsen, 1979; Clark, 1983; Kogan, 1984) remain largely unexplored. Given the magnitude of the demographic changes confronting higher education, there is clearly a need for theoretically based, empirical research focusing on how those changes are likely to affect postsecondary institutions. To provide an impetus for addressing that need, this chapter reviews demographic theory and research and explores implications of that work for higher education.

Necessarily, the review cannot and does not cover all aspects of demography in higher education. As noted earlier, we focus mainly on faculty demography, rather than on the demography of staff or students, and mainly on distributions by age, rank, and length of service. Questions concerning the changing gender and racial/ethnic distributions of faculty members are equally or perhaps more important, but have been considered often and well by other analysts (e.g., see Tolbert et al., 1995). Also, we focus here on the organizational effects of demographic distributions rather than the factors causing demographic change in higher education.[2] Furthermore, we focus more on academic units than on whole institutions or systems. At the level of programs, departments, and colleges within institutions, some of the most vital work of institutions takes place. It is there that teaching, research, and service are based, there that strategic planning primarily focuses, there that salary, tenure, and promotion disputes first arise, and there that accreditation bodies most frequently concentrate their attention (Peterson, 1976). Finally, we focus on academic units in universities rather than those in other kinds of institutions. Higher education is a diverse enterprise, and seniority distributions within higher education are similarly diverse. We believe the university sector, with its faculty predominantly full-time, tenure-track, and working in relatively discrete academic units, faces some of the most difficult of contemporary demographic challenges, and therefore is a fitting focus for this review.

In summary, we consider here the potential organizational effects of faculty seniority distributions within universities' academic units. There are many other ways we might defensibly investigate demographic issues in higher education, but this particular focus is not only significant from a policy perspective but also in keeping with a long research tradition in the field. Studying faculty in academic units has a history long enough, a literature robust enough, and analytic challenges difficult enough to warrant our choice of focus.[3]

---

[2]For an example of this provocative stream of research, see Tolbert and Oberfield (1991).

[3]See Dressel, Johnson, and Marcus (1970); Blau (1973);Smart and Elton (1975); Peterson (1976); Ramsey and Dodge (1983); Clark (1983, 1987); Baird (1986, 1990); Whetten and Bettenhausen (1987).

SENIORITY DISTRIBUTIONS IN THE DISTINCTIVE CONTEXT OF HIGHER-EDUCATION INSTITUTIONS

Much of the literature on organizational demography can be translated directly to higher-education institutions. Nevertheless, some distinctive features of colleges and universities may affect the influences of such demographic factors as seniority distributions. We review some of these features below, recognizing that not all of them are present or salient in every sector of higher education and that some are paralleled in other kinds of organizations. Yet they do stand as potentially important mediators of the relationships between demography and organizational structure and functioning in higher education. This section catalogues special features of higher-education organizations in three areas (the academic career, the academic task environment, and institutional organization) and provides examples of the kinds of connections these features might have to seniority distributions.

The Academic Career

A typical faculty career is quite different from careers in business, government, or even elementary/secondary school settings. Among the distinctive aspects of careers in higher education are the nature of academic preparation, tenure, advancement, and retirement.

In regard to *preparation*, the vast majority of academics have graduate degrees, either masters or doctorates. The years required for such advanced education virtually guarantee that the lower end of the age range among faculty will be above that of many other occupations. Doctoral work can and often does stretch to a decade or longer (Bowen and Rudenstine, 1992). Moreover, in applied fields such as education or engineering, newcomers must first acquire experience in the field before entering the faculty ranks. The long process of scholarly preparation serves to contract demographic differences in two ways. First, newly appointed faculty are older than neophytes in some other professions. Second, there are pronounced similarities in the experiences and preparation of graduate students within an academic discipline, no matter where or when they attend graduate school, which provide at least some common experiential ground for younger and older faculty.

Colleges and universities make many kinds of academic appointments, but in most institutions the core faculty are found in *tenured or tenure-track positions*. Among these faculty, nothing distinguishes junior from senior faculty more clearly than tenure status.[4] The goals, agendas, and priorities of probationary (untenured) faculty are strongly influenced by the tenure process's usual demand that faculty failing to achieve tenure after a set period leave (the "up-or-out" rule). Thus, turnover can be high among junior faculty. In contrast, among tenured faculty, the protection offered by tenure can restrain faculty mobility (Burke,

[4]As noted earlier, tenure status is like rank in being an aspect of seniority, and therefore might be considered demographic, in the sense that we use the term here.

1991). While some senior faculty move readily from one institution to another, many faculty remain at the institution that first offers them tenure, in part because of the mutual investment they and the institution have made during the probationary period. Few tenured faculty change institutions without a guarantee of indefinite tenure. What is more, it is difficult to dismiss a tenured faculty member. The tenure system, therefore, has been criticized as decoupling academic employment from evaluation processes (Hellweg and Churchman, 1981). In summary, turnover in higher education can be high at the junior levels and constrained at the senior levels.

In an academic career, the typical *advancement* ladder is very short: assistant professors become associate professors, who become full professors. Administrative appointments, endowed chairs, or specially distinguished appointments may add to a professor's title and responsibilities, but no higher rank exists. Rank is an important reflection of academic experience, as a promotion is conferred only after extensive examination of a professor's work and thus can be a useful indicator of power in academic institutions (Pfeffer and Moore, 1980). The three-step structure, together with what is usually a timely advancement process, ensures that many faculty spend most of their careers in the full-professor rank. In contrast to business or manufacturing concerns, there is no formal requirement in higher education for lower-ranking line personnel to pursue distinctively lower-ranking job responsibilities. Without a significant and continual infusion of new faculty, then, departments naturally and rapidly become demographically top-heavy when measured by rank alone.[5]

Academic appointments allow more naturally for continuing institutional affiliation after retirement than other careers. Academic work is not usually as physically demanding as manual labor, and thus a decline in health may be less relevant to employment. A faculty member near retirement may actually be at a professional peak (Clark and Lewis, 1988), whereas in business he or she might have been dismissed before ever coming close to traditional retirement age. Also, when faculty research is largely independent of university support and oversight, faculty can maintain their scholarly agendas into retirement, or at least well beyond the point at which an employer in another sector might terminate most employees. Retired faculty may also serve as valuable resources for particular teaching or advising tasks. Thus, in some cases, a department's demographic profile must take into account the continuing participation, contribution, and influence of retirees.

The Academic Task Environment

The nature of scholarly work can also influence the relationship between demography and other organizational variables. Academic tasks themselves, as well as

[5]This logic suggests that rank might advisedly be combined with or weighted by years of service or some other time-dependent variable to distinguish newly promoted full professors from those with much longer careers (Anderson, 1990).

the traditions of autonomy and self-regulation, affect interactions among faculty and assumptions about how the institution's work should be accomplished. The academy's distinctive patterns of exchange with external organizations also have implications for demographic effects.

Much of faculty members' *scholarly work* centers on the basic tasks of knowledge generation and transmission (Clark, 1983). Faculty recruits are usually selected by departments for their distinctive and complementary expertise (Burke, 1991). In research universities, faculty pursue their own research projects and teach within their own specialties, with some contributions to the department's basic courses. The work of the typical university department is little oriented to the *joint* production of knowledge or instruction. Rather, prestige and rewards accrue to faculty who are successful in their disparate research ventures while maintaining at least a modest commitment to the university's instructional goals (Lewis and Becker, 1979). Under these circumstances, the implicit hierarchy of experience and authority associated with a particular demographic distribution has little relevance to the goals or everyday work of professors. Junior and senior faculty alike are often better connected to scholars in their disciplinary specialties at other institutions than to their local colleagues (Becher, 1989). In many fields, senior faculty have little direct influence on the work of their junior colleagues, and junior colleagues may surpass some of the older group in scholarly production. Neither rewards nor authority in scholarly work, therefore, corresponds strictly to demographic status.

Faculty generally enjoy relatively high levels of *autonomy and self-regulation* in their work. Faculty have considerable control over their agendas, schedules, and workloads, subject to episodic review by other faculty or administrators (Clark, 1983). Despite accountability pressures and the existence of a formal hierarchy of administrators, the primary control mechanism is peer review for appointments, promotions, publications, funding, awards, disciplinary action, and censure (Braxton, 1986). Junior faculty generally have almost as much autonomy as senior faculty, but that freedom is balanced by evaluations by senior colleagues at critical junctures, such as tenure and promotion reviews (Tierney and Rhoads, 1993). Junior faculty who stray too far from their fields' established directions and methodologies in research risk losing the valuable support and endorsement of senior faculty.

Institution-environment relations are also rather distinctive in higher education. From a demographic perspective, it is the relationship between individuals and environments that is critical. Unlike enterprises in other sectors where designated individuals or units manage external relations, universities exhibit high levels of environmental attention and exchange at all levels and in virtually all units (Hearn and Heydinger, 1985). Most faculty maintain their own complex networks of interactions with external colleagues, funders, and organizations. The faculty most successful in securing external funding develop largely independent research enterprises, sometimes leading to entrepreneurial spin-off ventures (Louis, Blumenthal, Gluck, and Stoto, 1989). Short of that extreme, many faculty supplement their university employment with external consulting (Boyer and Lewis, 1985). Employment conditions and the mutual responsibilities of faculty and their institutions are complicated by these aspects of academic work. External relations yield, among other things, contacts, resources, and information, all of which can affect an individual's or group's position in the power structure of a department.

If individuals in different cohorts have different kinds of external connections, then there are demographic implications. As an example, consider the case of faculty turnover. When new faculty arrive, they usually introduce new external ties to their department, with consequences for the way the department functions. When faculty retire or resign or are denied tenure, some environmental relations are lost. At the aggregate level, junior and senior faculty may tend to differ in their individual environmental predilections. For example, junior faculty in an engineering department may prefer to pursue national sources for research funding, while senior faculty with local ties and experience may tend to prefer working with regionally-based engineering firms.

Thus, transitions toward one or the other end of the seniority distribution may imply changes in institution/environment relations. Of course, in most organizational settings, individual staff can make a difference in external relations—this is the essence of sales efforts, for example. But in higher education, the distributional tendencies in individual faculty members' relations with external agencies can profoundly influence the core "business" of the overall organization.

Institutional Organization

The organizational form of colleges and universities usually conforms to the general pattern of departments arrayed under a central administration. Both vertical and horizontal dimensions of this arrangement have implications for the study of demography.

Regarding the *vertical* dimension, there can be little question that universities have notably flat organizational structures (Blau, 1973). Though the central administration may have many levels and may be part of a broader system-level structure, the hierarchy that matters to most faculty in their everyday lives is that of the academic department. Here the formal hierarchy is usually minimal: one chair or head and, perhaps, an associate chair or two with limited responsibilities. Though some chairs are quite powerful, the tradition of *primus inter pares* is often maintained. In this case, the chair position rotates among the department's senior professors and provides only temporary, highly circumscribed authority. Real authority usually resides in the faculty themselves and is expressed through the plethora of committees that develop initiatives and make decisions. In consequence, power and influence often correspond less to administrative role than to experience, personal attributes, prestige, and savvy. Under these circumstances, demographic distinctions may be useful proxies for the faculty characteristics that determine critical, if implicit, hierarchies within departments.

The *horizontal* dimension is highly elaborated in higher education, with academic departments serving as the primary structural blocks on which most universities are built (Kerr, 1972; Clark, 1983; Coleman, 1973; Alpert, 1985). Sometimes, departments are supplemented by free-standing or inter-departmental research centers, yet most full-time, tenured or tenure-track faculty have primary appointments in departments. For the purposes of demographic analyses of organizational issues, institution-level demographic distributions are little more useful than broad national or state-level demographic distributions of faculty. Attention to the academic department appears to offer greater potential, for it is in the department that demographically distinct groups most often meet and clash or mesh.

Concluding Observations—Distinctiveness and Similarity

Clearly, academic organizations are not entirely like other organizations, and these differences imply a need for care in translating findings from more generic organizational-demography research into these settings. At the same time, there are similarities worth remembering. Colleges and universities are not immune to the universal patterns and forces of organizational life: a mission is stated in written, albeit often vague, terms; hierarchy among employees is formally recognized; there is a clear division of labor with horizontal and vertical differentiation; norms and values emerge and socialization occurs; various kinds of inequalities are noticed and important; conflicts occur; procedures and standards are established and reformed; some people move up in the organization more quickly than others, largely on the basis of expertise; turnover is endemic; authority relations exist; problems are solved in patterned ways; and so forth. Although research and theory from other settings must be employed cautiously, that research and theory can still be useful in aiding our understanding of academe.

THREE ARENAS FOR RESEARCH ON DEMOGRAPHIC EFFECTS IN UNIVERSITIES

In this review, we consider three kinds of effects that faculty seniority distributions might have on academic organizations: effects on academic structure, on organizational processes, and on interpersonal relations in the organization. Figure 1 presents these general categories as well as various sub-categories that appear in the current literature as influenced by patterns and transitions in organizational demography. We review the existing literature on demographic effects in these three domains and suggest potentially fruitful research initiatives associated with each.

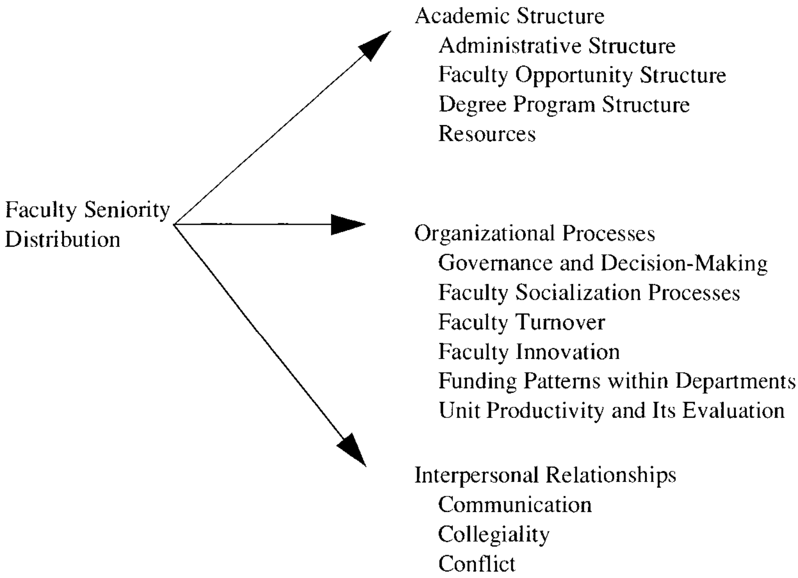


FIGURE 1. Effects of Faculty Seniority Distribution in Academic Units

Demographic Effects on Structure

The seniority distribution of faculty in an academic department has potentially profound effects on the department's structure. The effects of a department's seniority profile on the organization of its people, roles, authority framework, and even physical space provide a variety of promising avenues for research. Here, we focus on four aspects of structure that may be influenced by demographic factors: administrative structure, faculty opportunity structure, degree-program structure, and resources.

Administrative Structure: A department's administrative structure includes its appointed administrators, its committees and other support systems, and its formal rules and procedures. Perhaps the most intriguing link between demography and administration lies in the choice and role of the department chair or head. Zajac and Westphal (1996) present evidence from the corporate sector that, in choosing a CEO from outside the corporation, boards of directors tend to select someone demographically different from the previous CEO but more demographically similar to themselves.[6] That is, leader succession is viewed as an opportunity for change, but demographic similarity appears to reduce the ambiguity and uncertainty inherent in the selection of an outsider. In academic depart-

[6]Also see Westphal and Zajac (1995).

ments, faculty may exhibit the same tendency toward "self-cloning," especially in cases when a chair is recruited externally. Demographic similarity is likely to be defined with reference to the most powerful demographic group in a department, generally but not always the senior faculty.

A chair faces very different challenges depending on whether the department is predominantly senior or junior. The unusual, but not unheard-of, case of a relatively junior chair further complicates the chair's relationship to the faculty. There are instances in which associate professors have been appointed chair when irreconcilable conflicts among the senior faculty precluded a majority vote for any one of them, or when all members of the senior faculty group have served previously as chair and none could be persuaded to take on the task again. The demographic group from which the chair is chosen may have symbolic significance. The appointment of a senior successor to a senior chair may signal a continued concentration of authority in the group with longest tenure, while the first appointment of a chair from a mid-level faculty cohort (e.g., perhaps a recently promoted full professor) may represent the coming of age and emerging power of that faculty subgroup.

The rest of the administrative structure of a department may also be related to demography. Pfeffer (1983) has suggested that an organization with a large group of senior members may be able to function well with only minimal administrative structure because of widely shared assumptions and expectations about the department's work and processes. An influx of new faculty might necessitate greater administrative complexity as new ventures are added without older tasks being abandoned. A large group of junior faculty might also initiate a re-examination of departmental goals or a reorganization of responsibility, usually with some increase in the number of *ad hoc* committees. Competing groups of older and younger faculty might complicate both the appointment and the independence of committees, as faculty become sensitive to the ability of committees to represent their interests. Perhaps the most extreme demographic effect on administrative structure occurs as a dominant senior cohort approaches retirement and, over a period of several years, transfers its roles, responsibilities, and authority to a younger and less experienced cohort.

Procedures can also become more formalized in response to demographic change. When rules and departmental history are lodged in the memories of senior faculty, there is little need to develop formal procedures for departmental business, but this approach breaks down when other faculty cohorts grow in size. Administrative issues that might earlier have been handled by a group of faculty talking casually in the chair's office might start to generate memos, regulations, and handbooks. These demographic effects on structure may have, in turn, second-order influences. Hage and Aiken (1970) suggest that increases in formalization and centralization can lead to decreases in innovation in an organization.

Faculty Opportunity Structure: Despite the relatively flat organizational structure of academic departments, there is still considerable role differentiation among faculty. Both formally and informally, faculty take on departmental tasks and responsibilities that reflect their particular interests and expertise. Senior faculty may tend to rely on past experience in deciding who should handle which matters; in a sense, the pairing of people and problems may become institutionalized over time (Cohen and March, 1986). Younger groups of faculty may find that discovering their own and others' individual talents to address particular kinds of issues is as critical as dealing with the problems themselves.

Differentiation among faculty is evident in, for example, administrative appointments, role assignments, salaries, and tenure status. When the chair position rotates frequently among senior faculty, most professors who aspire to the position will eventually attain it; when the chair is appointed for a seemingly indefinite term, there is usually no struggle for the job. In the intermediate case of short-term but renewable chair appointments, there may be considerable competition for the top job. When departments are allowed to elect their chair, rather than accept the appointment of a faculty member by a higher-level administrator, demographic differences can be translated into formal political competition.

When the senior group is large or when a large cohort of mid-career faculty become candidates for formal leadership, positions of influence and responsibility simply may not be available for all faculty who seek them (Wallace, 1988; Baron, Davis-Blake, and Bielby, 1986; Riley, Foner, and Waring, 1988). Other than the chair, there is an array of formal but often uncompensated positions available in most departments, including assistant or associate chairs, directors of graduate and undergraduate studies, and academic program heads. Such positions may be as often occupied by mid-level faculty as by senior faculty. Importantly, the informal nature of some role assignments in academe (such as undergraduate mentor, government contact, or maintainer of alumni contacts) may ensure that some roles are chosen and maintained by the same individual faculty members over many years. Thus, there is no necessary association between leadership positions and seniority.

Among the clearest differentiators of status among faculty are salary and tenure status, both of which are susceptible to demographic influence. Pfeffer and Langton (1988) found demographic homogeneity to be associated with more equal salary distributions. In an analysis of broad population cohorts, Berger (1984) found that people in larger cohorts are more likely to experience slower growth in earnings and generally lower levels of earnings. In higher education, imbalances among faculty cohorts can affect salary structures in a variety of ways. A relatively large junior cohort is generally less expensive for an institution, due to the lower salaries paid to newer faculty, and thus might make overall increases in salary levels possible. In contrast, junior faculty in a department dominated by senior faculty might find most available salary increases going to their senior colleagues, especially if such increases are awarded in proportion to current salaries. The exception to this case is the department that finds it necessary to compete nationally in attracting or retaining the top junior faculty. Highly qualified new recruits may command salaries and benefits that rival those of more senior colleagues. This phenomenon of "salary compression" can cause problematic imbalances in the department's salary structure.

In the same way, the tenure structure may also work to a junior faculty member's advantage or disadvantage, depending on demographic realities. In institutions that operate formally or informally with tenure quotas, a demographically senior department may be "tenured up," effectively precluding the possibility of a young faculty member's advancement to tenure. Alternatively, a largely senior faculty that is ripe for rejuvenation might promote its junior faculty inappropriately. They might, for example, be over-eager for the assistant professors to replace retiring faculty and therefore burden junior faculty with administrative assignments. They might also evaluate tenure candidates less critically than the university's central promotion and tenure committees, whose approval the candidate must obtain (Burke, 1986), setting candidates up for failure. Research by Ferris, Judge, Chachere, and Liden (1991) provides support for a dissimilarity model of evaluation representing this phenomenon. They found that, among hospital nurses, supervisors who are dissimilar in age from the group they evaluate tend to give higher performance ratings than age-similar supervisors. The benefits to a junior faculty member of a senior-dominated department are affected by the eminence of the senior professors: successful senior faculty might provide a wealth of opportunities, contacts, and advice to the tenure candidate, but professors who themselves would have difficulty satisfying new requirements for tenure might not contribute to the candidate's success by either their advice or their endorsement.

Degree Program Structure: Undergraduate and graduate degree programs reflect both disciplinary and institutional orientations. Commonalities in content, process, and trends are evident across most departments within a discipline; however, individual degree programs also reflect the particular curricular and research histories of their departments. Demographic change within a department can influence degree structures through both disciplinary and institutional mechanisms. Young faculty may introduce the new perspectives or approaches of the discipline (McNeil and Thompson, 1971; Messeri, 1988), but their influence interacts with the local department's instructional legacy. While disciplinary or pedagogical innovation may eventually replace earlier models (Kuhn, 1970), the usual route to change in degree programs is through accretion, i.e., additions to the current structure (Kerr, 1972). This logic suggests that students in a department with a large group of junior faculty may face relatively complex degree requirements. They may also find that the departmental reputation (for either quality or some special disciplinary emphasis) that attracted them to that particular institution is based on the careers of faculty who are near or already in retirement.

Resources: Like salary patterns, the distribution of resources (including funding, support staff, and materials) can be indicators of faculty status and thus subject to competition (Deutsch, 1973). For example, resources contribute to what Merton (1968) has termed the "Matthew Effect," whereby past successes of faculty contribute to further advantages. Senior faculty whose grants or contracts pay for secretarial support or graduate assistants can marshal these personnel to out-compete younger, less-endowed faculty. Demographic divisions might therefore intensify divisions in academic social class.

As an indirect but important determinant of the distribution of resources in a department, seniority distributions might influence the allocation of physical space. A department full of senior faculty with rather large retinues of office or laboratory assistants requires more space than a department of junior faculty who have less grant support. Seniority effects on resources are mediated, however, by faculty productivity. For example, less active faculty whose earlier work earned them comfortable accommodations have sometimes lost those resources to new recruits who either bring external funding with them or negotiate contracts with substantial set-up packages.

Demographic Effects on Organizational Processes

Seniority distributions unquestionably affect a variety of processes at the organizational level, including unit governance and decision making; faculty socialization processes; faculty turnover; faculty innovation; funding patterns within units; and unit productivity and its evaluation. These aspects will be addressed in turn.

Governance and Decision-making: Several theorists have addressed the decision-making implications of rank distributions in organizations in general. Pfeffer (1981b) noted that the seniority distribution of organizational members is an important variable linked to change, adaptation, and leadership succession. Hage (1980) has suggested that, as the number of high ranking employees increases, pressures rise for more say in determining policies, budgets, and programs (i.e., greater decentralization in power), and value preferences become the main basis of group formation and decision making. In contrast, as the numbers of low ranking employees increase, power becomes more centralized and hierarchical levels and social characteristics become the prime bases of group formation and decision making.

As noted earlier, rank is not strictly connected to the availability of positions in higher education. That is, unlike, say, a bank, an academic department could be totally comprised of people at the highest possible rank with none at the lowest rank. In a sense, there is no need for "tellers" in academic departments—a solid cadre of "vice presidents" will suffice. Clearly, decision processes will vary depending on the extent to which there is an absence of lower-level personnel.

High faculty rank is associated with greater influence in decisions concerning appointments and curriculum (Blau, 1973). Indeed, many decisions in such settings may have been essentially "made" in earlier years, and little new discussion is required or encouraged (Corwin, 1969).

But decision making may be more problematic in contexts of demographic heterogeneity rather than dominance. Tsui and Ashford (1991), in a study in the corporate world, contrasted two theories regarding decision making in contexts of demographic heterogeneity. According to similarity-attraction theory, demographic differences should decrease feedback-seeking by managers ("These folks are different, so I will avoid them..."). According to the adaptive self-regulation perspective, demographic differences should increase such feedback-seeking ("These folks are different, so I need to know them better..."). Tsui and Ashford found that corporate managers they studied did not tend to seek feedback from demographically different peers, supporting the similarity-attraction propositions. Such findings are in keeping with the conclusions of Pfeffer (1983) and Davis-Blake (1992) that, when there is no clear dominance in departmental rank or seniority distributions (i.e., when the rank distributions are quite heterogeneous), the result may be a less integrated, less governable unit.[7]

This proposition may require qualification, however. Work by Massy, Wilger, and Colbeck (1994) suggests that the effects of heterogeneity in faculty seniority distributions tend to vary by institutional and departmental goals. These analysts found that heterogeneity tends to be associated with more strained decision making in research-oriented institutions, where senior faculty told the researchers that their junior colleagues were "difficult" and held too lofty opinions of themselves. Within teaching-oriented settings, however, status differences were de-emphasized and relationships across seniority levels were more collegial. In those settings, senior faculty viewed younger faculty as representing the future of their departments and appreciated their ideas. Communication between younger and older faculty was more frequent, and senior faculty involved junior faculty more directly in major decisions. Senior faculty expressed admiration for assistant professors, junior faculty viewed senior members as mentors, and junior and senior faculty reviewed each other's work. In summary, the work of Massy and his colleagues emphasizes the possibility of interactive effects of institutional type and demography on governance and decision making.

Over time, departmental planning and decision making may benefit appreciably from the presence of junior faculty. Karl Weick (1977) has commented on the paralyzing penalties that organizations pay when they allow dominant, and perhaps ineffective or misguided, views of external environments to lie unchallenged for long periods of time. In support of the proposition that long-standing dominance by senior personnel may have deleterious effects, Bantel

[7] Support for this notion from outside higher education has come from Smith et al.'s (1994) work on top-management teams.

(1994) found that executive teams with long years of tenure in an organization tended to resist analysis of changes in external environments of their firms. Similarly, Wiersema and Bantel (1992) found that seniority distributions may contribute to stagnation in large, diversified corporations. Specifically, they found that lower average age and shorter organizational tenure among staff tend to be related to substantive changes in organizational direction. These authors also found that individuals take on the cognitive perspective of the dominant, more senior members of a team, and thus become fully assimilated, only after about eleven years.

Intragroup heterogeneity in terms of organizational tenure has been found in other contexts to have both positive and negative effects on group performance. In a study of high-technology companies, Ancona and Caldwell (1992) found that greater tenure diversity tended to improve a team's ability to clarify group goals and to set priorities, which led to higher team ratings of performance. This outcome came, however, at the cost of a greater need for the team to develop negotiation and conflict-resolution skills.

In a similar analysis, Finkelstein and Hambrick (1990) found that longer average tenure of executive teams was associated with corporate strategies that tended to conform to dominant tendencies in firms' home industries, and that those strategies tended to lead to modal rather than extreme results in financial performance. In contrast, these analysts found that less-senior executive teams exhibited more unusual strategies and more dramatic results, in both positive and negative directions. The tendency of senior leadership to resist change was also exhibited in a study by Kosnik (1990) which suggests that corporate trustees dominated by members with longstanding board involvement were more likely to resist "greenmail"-inspired buyback schemes and other innovative financing tactics. Change and new directions are not necessarily good for organizations, of course, but this line of research at least suggests some possible benefits of seniority heterogeneity for effectiveness in higher education.

Faculty Socialization Processes: The ways in which newcomers are treated in organizations are fundamental to organizational culture, adaptation, and survival. Several studies have focused on the nature of socialization efforts in demographically diverse units. On the basis of research in the corporate sector, Riley, Foner, and Waring (1988) suggest that units characterized by vast age differences may experience controversy over socialization. For example, communicating and reproducing the nature of the organization may be problematic when there are fundamental, demographically linked differences in the peoples' perceptions of the organization.

A 1971 study by McNeil and Thompson, one of the most thorough analyses of academic demography to date, pays close attention to faculty socialization as a facet of the "demographic metabolism" of university departments. These authors focus on the "social regeneration" of organizations, i.e., patterns over time in the ratio of newcomers to veteran members. They argue that a healthy rate of regeneration, neither too high nor too low, implies that a sense of stability prevails in an organization, even though people enter and leave. Obviously, McNeil and Thompson argue, the socialization process tends to be easier during slow and steady regeneration. If an organization is stable, more attention can be paid to meeting its fundamental goals. As regeneration rates go higher, however, the proportion of members who know the organizational culture declines, and the need for socialization and routinization necessarily increases. A smaller proportion of veterans must socialize newcomers and, although veterans have leverage, new customs and norms are likely to develop among the ascendant pool of newcomers. Thus, when units are forced to recruit heavily among the younger population —that is, to pursue rapid regeneration—departmental turmoil may result.

Others have echoed McNeil and Thompson's arguments, observing that the nature of age, rank, and length-of-service distributions in departments can affect the ways in which junior faculty are socialized into academic life. Work by Corwin (1969) suggests that older, homogeneous, and relatively stable faculties may be especially willing to enforce strong socialization regimes and expel non-conforming new members. Also in keeping with the McNeil and Thompson perspective is work by McCain, O'Reilly, and Pfeffer (1983) arguing that people who enter an organization together or at about the same time are more likely to associate with each other and to have similar perspectives on the organization and its operations.

More recent research has focused on connections between the intensity of socialization efforts and the rates of turnover and promotion within organizations. Of particular relevance to higher education is research on organizations with "up or out" requirements. According to a recent review by Malos and Campion (1995), junior members of such organizations may be less likely to ascend to seniority when little time is spent on training, mentoring, and providing high-quality work assignments. Similarly, effectiveness can sometimes be compromised when new hires are at the senior rather than at the junior level. The absence of the benefits of distinctively localized socialization while faculty are at the junior level may inhibit smooth departmental functioning.[8] Thus, units tilted to the upper ranges in seniority may benefit from a commitment to bringing newer, junior members into the fold, training them appropriately, and promoting them to senior status. Historically, departments in some elite universities in this country have been noted for their tendency to hire senior faculty from other institutions rather than promoting junior faculty from within. The work of Malos and Campion suggest some of the penalties possibly inherent in such an approach.

Faculty Turnover: Pfeffer (1981b) has stressed that organizational age and

---

[8]Long ago, Kingsley Davis (1940) observed that socialization processes are more problematic among older people, who tend to cling to basic values and attitudes acquired and reinforced in earlier life.

length-of-service distributions can have notable effects on organizational turnover.[9] In particular, there is ample evidence that employees' length of tenure in the organization is negatively related to departure: people with more years of service in an organization are usually more satisfied and tend to remain there. The odds of their leaving in a given year are low. In higher education, the pattern is the same: Smart (1990) has found that, regardless of tenure status, younger faculty are more likely to leave their institutions.

These individual-level findings parallel organization-level findings that, to the extent an organization is demographically heterogeneous, particularly in terms of seniority and age, its participants will experience less social integration and poorer communication, problems that can lead to turnover (see Pfeffer, 1985; O'Reilly et al., 1989; Jackson et al., 1991; Davis-Blake, 1992; Wiersema and Bird, 1993). In a particularly interesting study in this genre, McCain, O'Reilly, and Pfeffer (1983) found that university academic departments with substantial gaps among cohorts were characterized by increased rates of voluntary retirements, resignations, involuntary removal, and expired appointments. As noted earlier, their explanation for the finding was that people who enter an organization together or at about the same time tend to view the organization similarly. Identifiable discontinuities or bulges in the organization's demographic distributions thus can create ripe conditions for turnover. Pfeffer and O'Reilly (1987) also found that, among hospital nurses, diversity in terms of tenure in the organization was positively related to higher levels of turnover.[10]

Several studies at the unit level have focused on turnover in leadership ranks, rather than employee turnover in general. Wagner, Pfeffer, and O'Reilly (1984), for example, found in a corporate study that date-of-entry distributions predicted the proportion of a top-management group that left an organization. Specifically, the more disparate the dates of entry, the higher the level of turnover experienced in the leadership group. Interestingly, seniority distributions may interact with other factors in a department to affect unit heads' stay in their positions. Pfeffer and Moore (1980) found that an academic discipline's paradigm state (i.e., the level of development, solidarity, and codification of the discipline) works in conjunction with demographic factors to influence the longevity of an academic department chair's stay in that position. Within fields in highly developed paradigms, a department's level of faculty seniority was positively related to the department chair's time in office. Within fields with less developed paradigms, however, the unit's level of faculty seniority was negatively related to the

---

[9] More recently, Milliken and Martins (1996) have echoed this conclusion.

[10] Research by Tolbert, Simons, Andrews, and Rhee (1995) suggests a similar effect of gender-related demography on turnover: as a department's proportion of women increased, turnover also increased, at least until the proportion of women reached the 35 to 40 percent range. Such a finding regarding gender-composition "tipping points" supports competition theory rather than social-contact theory (see Kanter, 1977). That is, the growth of a minority group from token status led to increased rather than decreased intergroup competition and conflict.

department head's tenure.[11] Seniority could conceivably provide more contenders for a position and thus more turnover, or seniority could suggest more socialization into a common culture and thus less turnover. It appears the first explanation holds in less developed fields and the second applies to more developed fields. Pfeffer and Moore suggest that, in and of themselves, faculty seniority and scholarly consensus may not be powerful predictors of chair turnover. Together, however, they can make a difference. Results like these indicate that scholarly consensus among senior faculty, in particular, may be critical to administrative stability.

Faculty Innovation: Demographic distributions may also affect rates of innovation in teaching, research, and service. Hage and Aiken (1970) have hypothesized that increases in stratification in organizations may deter innovation. One could argue that increasing demographic diversity in rank, age, and seniority is by definition increasing stratification. For example, the addition of a junior faculty member to a previously all-senior department marks a new kind of hierarchy in that unit. Thus, by Hage and Aiken's conception, innovation in the department might decline. Support for this perspective comes from research by Zajac, Golden, and Shortell (1991), who found that organizational innovation among corporations in the health professions was greatest when the age distribution of organizational members was homogenous (i.e., regardless of whether that distribution was on the older or younger end, or somewhere in the middle). On the other hand, conventional wisdom seems to suggest that new faculty can bring new ideas.

The two views are not necessarily mutually exclusive. For one thing, innovative thinking at the individual level may not translate into organization-wide innovation. At the departmental level, questions of individual research approaches are less relevant, and less visible, than questions of how departmental business is accomplished. More fundamentally, organizations may discourage the hiring of new professors who might "rock the boat", and those new colleagues who threaten to do so may be quickly socialized and sanctioned away from innovative behavior.[12]

Overall, there is no clear answer to the question of the relationship of seniority distributions to organizational innovation. Pfeffer (1981b, p. 15) suggested that, as of the early 1980s, the connections between demography and innovation were largely untested. More research has now been conducted on the topic, but evidence of a direct causal relationship is still lacking (Milliken and Martins, 1996).

Funding Patterns within Units: Younger and more junior faculty tend to fare differently from their more senior counterparts in the competition for

[11] For a review of this and related studies, see Braxton and Hargens (1996).

[12] In an intriguing historical study, Reed (1978) found that older cohorts in the American Foreign Service in the 1940s recruited and promoted younger cohorts who resembled themselves in career preparation and content. As a consequence, he suggests, the Service's policy positions remained largely unchanged as well, despite a significantly changed geopolitical context. He concluded (p. 418) that this "helped to maintain an elitist Officer Corps focused on traditional diplomatic functions."

research funding, and departments with a rather junior demographic composition overall may tend to have rather disparate patterns of funding. Over time, therefore, movement from senior to junior composition or vice-versa may affect the overall pattern of funding in a department. Anderson (1990) found that the accumulated years of experience within an academic department may affect the group's ability to shield itself from internal budget cuts and its capacity for mustering external research support. As departments become more senior, they become more dependent on a limited range of relatively critical sources of support. For example, in departments that were heavily dependent on federal research dollars at a given time, the proportion of funding from federal sources grew as departmental faculty increased in seniority. In contrast, when faculty in such units became more junior, the departments tended to shift away from traditional federal funding sources. Such shifts have differing implications for departments' ongoing financial arrangements and relationships over time.

Demographic factors may also heighten competition for resources among faculty in departments with demographic "bulges." Funding from either institutional or external sources is sometimes linked explicitly or implicitly to career stage; consider, for example, grants targeted at tenure-track faculty or, alternatively, grant programs that virtually require the track record of a successful senior faculty member. Faculty in large cohorts may find themselves competing frequently with each other for such funding.

The level of homogeneity in age or seniority may also affect the ease with which departments are able to allocate internal (institutional) funds. Riley, Foner, and Waring (1988) suggest that corporations with vast age differences tend to experience internal conflicts over budgeting, funding, and perceived inequalities among age strata. Although research on this topic is limited in higher education, the findings from other settings seem to suggest an intriguing proposition testable in departments with some decision-making authority over discretionary funds.[13]

Unit Productivity and Its Evaluation: Research on the connections between faculty age and productivity at the individual level may provide some guidance for demographic analysis at the organizational level. That is, such research may suggest the implications of "top-heaviness" or "bottom-heaviness" for departmental productivity. Studies at the individual level may be categorized as to their focus on research, teaching, or the full range of faculty work. Regarding research, most studies have shown that professors' productivity is not directly limited by age.[14] Notable in this group is the extensive review of the literature by the National Research Council (Hammond and Morgan, 1991). That review supported the finding of an earlier, similar review by Reskin (1979, p. 203) that "in

---

[13]Pfeffer and Langton's work (1993) might prove useful for future analyses of this kind in universities.

[14]Some analyses (see Levin and Stephan, 1989) have suggested otherwise, however.

no case did productivity show a simple negative relationship with age." That is not to say, however, that older and younger faculty do research in the same ways or with the same goals. In an intriguing analysis, Messeri (1988) noted that greater awareness of scientific development as well as motivational factors predispose young scientists to adopt scientific innovations before their more senior colleagues. Messeri notes, however, that older faculty are in some ways better suited to pursue unconventional research and to advocate controversial ideas, because experience and tenure may provide them with social and professional opportunities to offset the high costs and obstacles associated with risky research endeavors.

Regarding teaching, the evidence is similar. Although teaching styles and preferences vary somewhat by age, Baldwin (1979), Finkelstein (1984), Hammond and Morgan (1991), Kinney and Smith (1992), and others conclude that there is no clear evidence that teaching quality or effectiveness is affected by the age of faculty.

These findings specifically for research and teaching are echoed in efforts to assess the full range of faculty values, attitudes and activities. A study by Dressel, Johnson, and Marcus (1970) found that senior faculty members tended to be more oriented toward their university, while junior faculty members tended to orient themselves more toward their academic disciplines. Later, in a similar vein, Finkelstein (1984) and Bowen and Schuster (1986) concluded that the connections between age and overall faculty performance are not at all clearcut: older faculty may do different things, express different attitudes, and have different preferences (e.g., for teaching versus research), but are not necessarily more or less productive. Along the same lines, Lawrence and Blackburn (1988) report analyses indicating that differences in age do not relate to differences in effort in faculty cohorts, but do relate to differences in activities and attitudes: they found older professors gave more time to university service than younger faculty and were also somewhat more interested in teaching. Such findings suggest that Gouldner's (1957, 1958) distinction between "cosmopolitan" and "local" faculty seems to differentiate older and younger faculty cohorts: localism may increase with age.

There is a modest but growing set of relevant research findings from the corporate sector on the effects of seniority composition on the performance of work groups. Some of the research suggests that groups that are highly senior in composition may not serve an organization well. For example, Davis-Blake (1992) has found that work groups who had been together longer than five years undertake less communication with sources of information outside of the group and therefore do not perform as well. Smith et al. (1994) found that top-management teams with diverse levels of experience encountered conflict in decision making, requiring more coordination and monitoring by a CEO. Nevertheless, those authors found that top-management teams' tenure distributions had no effect on communication, social integration, or group performance. What is more, time (specifically, length of service together) can apparently serve to dampen the initial negative, stereotype-based interactions among work groups that are diverse in age, gender, and ethnicity (Van Dyne and Jehn, 1992).

In another study of groups' age distributions, seniority distributions, and communication efficacy, Zenger and Lawrence (1989) found that different aspects of demographic heterogeneity limited communications, depending on the kind of task at hand. In particular, age distributions had more influence than tenure distributions on technical communication inside engineering teams but less influence on communication with external parties. Katz (1982) found that teams with high longevity communicated less frequently among themselves, with people in other parts of the organization and with external parties. This result held independently of age and organizational tenure.

Thus, the corporate evidence is mixed and complex regarding the effects of age and seniority distributions on groups' work performance. It is important to remember that these findings apply to groups working on a joint problem, such as the development of a new product or a corporate strategy. For this reason, they are probably most applicable in higher education to faculty research teams or departmental committees. Academic departments in research universities, in particular, rarely work together as one group. Without any direct evidence, therefore, it is hard to surmise how these findings might be applied to entire academic departments.

Examining together all the above research on demography and performance, it seems that the better part of valor would be to treat with great caution surmises suggesting how faculty seniority distributions influence the performance of units in higher education. Barring unusual definitions of quality, change in seniority distributions seems unlikely to have a direct effect on organizational effectiveness in the curriculum, in research, or in service activities. Indirectly, however, such effects might occur as professors' modal priorities for their work lives change. For example, as a department becomes more junior, it might find its faculty more and more oriented to research efforts, less and less oriented to teaching, more and more oriented to academic discipline rather than the local institution, and so forth. The quality of work products themselves would not necessarily be altered at the individual or aggregate (departmental) level by such changes, but the quality of the overall "product mix" of the department could indeed be altered. That is, changes in product mix might be seen as reducing or raising quality, depending on the extent to which such changes are in line with the university's mission. For example, legislators might be disappointed by a state university's business school moving more and more to "basic" research and away from applied projects with corporate sponsorship, owing to a transition toward more junior faculty. Such disappointment might well be unrelated to the scholarly quality of the specific research being done.

Demographic Effects on Interpersonal Relationships

The seniority distribution in an academic department can affect not only the unit's structures and processes, but also the relationships among its faculty. Of course, interpersonal relationships are connected in important ways with structures and processes. In this section we highlight certain interpersonal issues embedded in the previous sections, as well as some separate issues not yet addressed. Our special focus here is on seniority distributions' effects on interpersonal communication, collegiality, and conflict.

Communication: Fundamental to much of the research on organizational demography is the assumption that demographic similarity is a salient factor in interpersonal interaction and, further, that such similarity enhances attraction and communication (Pfeffer, 1983; Tsui and Ashford, 1991). Communication benefits from common assumptions that develop from shared experiences; to the extent that these experiences are related to periods of academic or social history, they are reflected in demographic similarity. Shared experience may be based on similarities in background, on the initiation and integration of faculty members who enter a department in roughly the same cohort, or on the accumulated experience of faculty who have been in the same department for a long time.

Demographic similarity has been linked to frequency or intensity of communication and thereby to a number of organizational outcomes, such as social integration and turnover (Davis-Blake, 1992); cohesiveness (Coombs, 1992; Pfeffer, 1985); and early retirements, voluntary resignations, and involuntary removals (McCain, O'Reilly, and Pfeffer, 1983). Coombs (1992) has suggested that, because homogeneous groups tend to communicate more frequently among themselves, cohorts are more likely to develop complete communication networks. It appears, therefore, that dominance of a particular faculty group may be less relevant in demographic effects on relationships than the existence of identifiable faculty cohorts. For example, McCain, O'Reilly, and Pfeffer (1983) based their analysis of demographic effects on faculty turnover on gaps and bulges in departments' seniority distributions. They reasoned that faculty members' identification with a cohort, enhanced by frequent association and communication, is the key to understanding demographic effects. When cohort divisions are less pronounced (that is, when the demographic distribution is relatively even), demographic similarity is less relevant to communication flows.

Collegiality: Demography influences the extent to which faculty in a department are able to work together as colleagues (Goodman, 1962; Millett, 1962). Much of academic life, especially in research universities, centers on the work that faculty do independently or in association with faculty at other institutions (Becher, 1989). Thus, in some departments it is possible to find faculty who have very little to do with each other and, in fact, scarcely know one another. To the extent that demographic similarity increases professors' interaction with one another, it may improve their chances of cooperating as colleagues. We provide three illustrations.

First, demography may affect the extent of collaboration in research and teaching within a department. An entering cohort of faculty may embody the disciplinary or pedagogical paradigms dominating at the time these faculty attended graduate school (Kuhn, 1970). Collaboration within demographic cohorts may thus be easier, due to the shared assumptions, perspectives, and values associated with a given paradigm. There are instances of departments intentionally recruiting groups of new faculty to establish centers of research in a particular area or with a particular theoretical orientation.

Second, demography may affect a department's ability to cooperate on those rare ventures that require the entire faculty's participation. Some research or training grants may be dependent on a department's ability to demonstrate a critical mass of support or expertise. A more extreme example is the case of a department facing elimination or dispersion. Here demographic effects are difficult to predict. A dominant cohort might have the capacity for marshaling the consensus and cooperation necessary to withstand such threats. Alternatively, a more even distribution might represent greater flexibility and strategic strength in responding to extreme challenges.

Third, demography may contribute to faculty members' ability to move beyond collegiality to friendships. Similarity-attraction theory (Tsui and Ashford, 1991) applies to personal, as well as professional relationships. When a demographic cohort is strengthened by personal friendships, departmental matters may become more complicated. Professional issues may be affected by personal loyalties. Cohort solidarity might distort discussions, interpretations, and decision-making. External groups or individual outsiders might find it impossible to interact with certain faculty without involving an entire faculty network.

The effects of demography on collegiality may be mediated by other departmental characteristics, such as size and role. Massy, Wilger, and Colbeck (1994) found that in small departments, particularly those devoted primarily to teaching, faculty are more likely to be able to work collectively despite differences in seniority. Collegial relationships in such departments, Massy et al. note, are strengthened by patterns of consensus-building, shared power, frequent consultation, and a sense of collective responsibility.

Conflict: Conflict is a normal and powerful force in organizational life, and higher education is no exception (Baldridge, 1971). Most academic departments bear the scars of past battles. The remembered histories of those battles often emphasize the roles of individuals and personalities. It is intriguing, however, to consider the role played by demographic patterns or demographic change in sparking or sustaining such conflict. The potential connection between seniority distributions and conflict has long been noted by organizational theorists. Karl Mannheim (1952) devoted a chapter to it in his classic *Essays on the Sociology of Knowledge*, and Joseph Gusfield (1957, p. 323), following the conceptual lead of Mannheim, suggested that:

> Conflicts of power and policy between age-groups are a common feature of many organizational structures. Factories, churches, labor unions, and political parties often distribute power, prestige, and income along an age-grade hierarchy. The existence of 'old guard' and 'young Turks' is found in many areas of society....The existence of generational differences leads to divergent political and social styles and modes of thought which greatly influence the character of public issues. When two or more generations appear within the same organization we may consequently anticipate factional conflict.

Thus, one of the most interesting potential effects of demography is its influence on the level of conflict in organizations. For this essay, we consider seniority distributions' possible connections to three forms of conflict in academic departments: intra-cohort, inter-cohort, and inter-departmental.

Some theorists have claimed that organizational conflict is rooted mainly in competition for scarce resources (e.g., see Pfeffer, 1981a). Most higher-education institutions have experienced enough financial strain over the last quarter century to provide ample opportunities to support the claim that the scarcest resource is money. Bowen (1980) maintained, however, that the more important resources for academics are quality and prestige and that it is impossible to acquire too much of either. Regardless of whether the focus is on finances or on other, more abstract resources, there is no question that demographic "bulges" and other imbalances have the potential for heightening competition and conflict within departments.

This reasoning would seem to be most applicable to *intra-cohort conflict*. Faculty in a large departmental cohort whose interactions are driven by either real or imagined competition can experience considerable conflict, despite their surface homogeneity. Whether this effect is stronger in senior or junior cohorts is an open question: senior faculty could have years of struggle behind them to confirm their competitive views of one another, while junior faculty might not only be competing for their professional lives but also fighting new battles without the benefit of adversarial precedent (Corwin, 1969). Competition may also center on comparisons that faculty members make between themselves and others in their cohort. Age norms and role expectations may be more sharply defined in a relatively large cohort, and non-conformity may be a particular source of conflict within the cohort (Waring, 1975; Lawrence, 1987, 1988; Hage, 1980).

While *inter-cohort* conflict could also be framed in terms of competition, in this case it might be more useful to consider value conflicts as the driving force (Riley, Foner, and Waring, 1988). Pronounced disagreements stemming from differences in intellectual perspectives could perhaps be traced to differences in the eras in which faculty were trained, socialized, and admitted to the profession. Interestingly, Hammond and Morgan (1991, p. 31) note that research universities in the 1990s have more faculty in the higher and lower age ranges than other institutions, and consequently proportionately fewer in the middle ranges. Under the logic suggested here, there might be more likelihood of conflict in those settings.

Finally, we turn to *conflicts between departments*. McNeil and Thompson

(1971) suggest that rapid regeneration in academic departments and the resulting gaps between newcomers and old-timers can lead to inter-departmental conflict. Strains develop, they claim, between departments dominated by veterans and those with many new recruits. Though departments can sometimes be astonishingly insular, most of them rely on other departments for contributions to their own curricula, graduate programs, research ventures, and so on. When a department's *status quo* is upended by a large group of new faculty, tensions derived from value conflicts may spread beyond the department and may particularly affect relationships with departments that have not experienced much regeneration. A department dominated by a group of junior faculty may come to have a reputation for being inept, arrogant, intentionally or unintentionally non-conformist, or radical.

FURTHER CHALLENGES FOR RESEARCH ON DEMOGRAPHY IN HIGHER EDUCATION

In Figure 1 (page 245) and above, we have presented and discussed an orderly agenda for research on the effects of organizational demography in academic institutions. We have reviewed literature on three kinds of demographic effects. Because most of that literature has related to non-academic contexts, we have presented a variety of potentially fruitful specific directions for extending that research into higher education. In this part of the review, we suggest some necessary and complementary elements of a research agenda in this arena. These may be viewed as further challenges for those interested in pursuing analysis of academic demography.

Extension of Research to Other Demographic Characteristics Beyond Seniority: In this review, our emphasis has been on the time-related dimensions of demography such as age, length of service, and so on. As we noted earlier, other characteristics such as gender, race or ethnic identification, country or region of origin, past organizational affiliations, or even religion in some contexts might be fruitfully added to or combined with time-dependent characteristics in the development of demographic research in academic settings.

Development of Demography as a Concept: In simple terms, we have treated demography as a organization-level variable. The complication in this conceptualization is that demography is actually a distribution of values of some underlying, individual-level variable such as chronological age, career age, length of time in the organization, rank, or even a qualitative variable such as nature or extent of experience or familiarity with a particular issue. Demography itself can be conceptualized in terms of the underlying variable (or variables) in a number of different ways: for example, as a range or clustering of its values, as a pattern in the distribution of its values, or as a balance between extreme and central values in the distribution. To emphasize its dynamic properties (since even a stable population shifts demographically from year to year, as long as demography is linked to time-dependent variables), demography could be conceptualized as a flow in the underlying variable or as a pattern of transitions among stages. There are many possibilities for creative conceptualizations of demography as an organization-level variable. Perhaps some of the most interesting research will emerge from new ways of thinking about demography.

Attention to Precise Measurement of Demography: Some analyses of organizational demography have relied on rather inexact summary indicators of demographic distributions, but a number of studies have used sophisticated approaches to the measuring demography (e.g., Pfeffer and Moore, 1980; Tsui and O'Reilly, 1989; Wagner, Pfeffer, and O'Reilly, 1984; McCain, O'Reilly, and Pfeffer, 1983). Many interesting demographic indicators have not received the review and widespread application in higher education that they deserve, and little attention has been paid to creating new indicators.[15]

Theory Development Concerning Demography's Other Organizational Effects in Higher Education: We have focused on seniority distributions as an independent variable and have reviewed their effects on structure, processes, and relationships. We can immediately suggest three extensions. First, within the three effect arenas addressed here, there are additional demographic effects that have not yet been explored. For example, one might investigate the effects of demography on departmental mission or goals as an aspect of structure; on faculty commitment or participation in extension, consulting, or other forms of outreach as aspects of process; or on mentoring of students as an aspect of relationship. Second, there may be other important areas of effects that do not fit congenially within the three-way typology proposed here. Departmental culture, for example, overlaps in part with structure, process, and relationships, but includes broader dimensions as well. What are demography's effects on language, symbolism, interpretations of experience, and the sense of meaning and purpose shared by faculty in a department? Third, there is much work to be done in strengthening the theoretical connections between demography and the other organizational variables it affects. The complexity of conceiving of demography as an organization-level variable and the variance in the clarity with which it has been conceptualized in the literature have hindered the kind of theory development that is needed. There are, however, well developed theories in sociology, political science, economics, and the broader organizational literature that could be used to support a theory of demographic effects.[16]

Attention to the Interconnections Between the Effects of Demography and the Effects of Other Organizational Factors: The effects of demography are never so isolated and separable as this review might suggest. We conceive of any empirical study of organizational outcomes as having a multivariate design in which the seniority factors we emphasize are only one part of a larger model including

---

[15] See Hearn, Anderson, and Eck (1996) for some initial work in this direction.

[16] See, for example, Tsui and O'Reilly's (1989) application of similarity-attraction theory and Haveman's (1995) use of organizational-ecology theory.

many hypothesized causal elements. Included in those other elements would be such additional demographic factors as gender and race/ethnicity distributions, as well as unit size, the nature of the department's home discipline, labor-market conditions, national ranking, governance arrangements, funding patterns, enrollment and advising patterns, and the like. Also notable among the non-demographic factors influencing organizational outcomes in academic units are a unit's disciplinary culture (see Becher, 1989) as well as its discipline's "paradigm status" and categorization under various conceptual typologies (see Braxton and Hargens, 1996).

In multivariate models of this kind, one would have to think of demographic factors as having effects not only distinct from the other factors' effects but also interactive with those other factors. One of the most striking suggestions of interactive demographic effects has been Pfeffer and Moore's (1983) finding that faculty seniority distributions have effects on chair turnover only in conjunction with the state of "paradigm" development in departments. The organizational distinctions made by such analysts as Lodahl and Gordon (1972) and Biglan (1973) among departments with different disciplinary underpinnings are therefore potentially quite relevant: perhaps departments with an established formalized, consensually agreed-upon knowledge base (e.g., chemistry) are less likely to suffer demographically generated conflicts than units with less developed "paradigms" and more reliance on informal, interpersonal decision making concerning critical departmental issues. Surely many of the other effects of demography are similarly contingent.[17]

Consideration of Demography as a Dependent or Intervening Variable in Studies of Higher-Education Organization: Though we have limited our discussion to demography's effects, Pfeffer's (1983) pioneering review of the early organizational demography literature explored the antecedents of demography as well. The three main sources of variation in organizational demography that he identified (personnel practices, growth, and technology) are all influenced by the distinctive characteristics of higher-education systems and institutions, such as tenure, career preparation, the academic labor market, and the scholarly work environment. Explorations of the determinants of demography may prove valuable, but perhaps of even greater interest would be studies that treat demography as an intervening variable. Given that organizational studies of higher education have largely ignored faculty demography and given that demography has proved to be an important predictor of organizational phenomena in other contexts, it is likely that examining demography's intervening effects would add significantly to the explanation of variation in institution-level variables in higher education.

Longitudinal Analyses of Demography: As noted above, demography is by nature a dynamic organizational variable. It lends itself readily to examination

[17]The study by Massy, Wilger, and Colbeck (1994) provides further evidence of interaction effects in this kind of research.

of organizational change over time through relatively recent advances in longitudinal research methodologies such as hierarchical linear modeling (Bryk and Raudenbush, 1992), event-history analysis (Willett and Singer, 1991) and other forms of dynamic modeling (Tuma and Hannan, 1984).

Application of a Broader Array of Research Approaches: Virtually all of the demographic analyses that have appeared in the literature have treated demography in quantifiable terms. Absent are studies of demographic effects through qualitative analysis. We could learn much more about the connections between faculty demography and other organizational phenomena through intensive observations of faculty conflict, collaboration, and other interaction across and within demographic groups. It would be intriguing to find out to what extent faculty interpret their relations with colleagues and the activity within their departments in terms of demographic differences and similarities, or, to go even further, to what extent faculty members' very perceptions of demographic difference shape their relationships and actions.

Extension of Demographic Analyses to Other Sectors of Higher Education: In this review, we have focused largely on demographic effects in U.S. universities. Organizational analyses in other sectors of U.S. higher education and other national systems would likely suggest other roles for demography. In the U.S., patterns of retirements and new hires differ appreciably by institutional type (see Clark and Corcoran, 1987). For example, a National Academy of Sciences study (Hammond and Morgan, 1991) found that the attractiveness of employment in research universities in the U.S. tends to discourage early retirement in those settings, and thus has the potential to lead to increased institutional costs, decreased adaptability to change, and decreased attractiveness to new faculty. In contrast, a variety of factors may make four-year and two-year colleges in this country better able to bring in new faculty and more adaptable to emerging demographic circumstances. Abroad, national central planning is more prevalent than in the U.S. system, so demographic effects may be different in nature or in strength. With such differences in mind, anyone undertaking research on organizational demography should attend thoroughly to the localized conditions of different institutions.[18]

CONCLUSIONS

Jeffrey Pfeffer, echoing Harold Leavitt years before, has noted that,

> Because of the individualistic values of the society, we tend to think of things in individualistic terms. Thus, we manage careers of individuals; concern ourselves with individual needs, attitudes, and demographic characteristics; and concern ourselves with finding the best individuals and rewarding them individually. The difficulty of this approach is its neglect of the interdependence and relationships that are the essential,

[18]See Massy, Wilger, and Colbeck (1994) for an intriguing study of institutional differences in the effects of seniority on departmental functioning.

> indeed defining, characteristic of organizations. For instance, to predict performance or turnover on the basis of an individual's sex is not likely to be productive, though knowing what the sex composition is of the organization or work unit does enable one to make some predictions about performance pressure and social integration. Similarly, tenure in the organization helps to explain turnover, but not as much as when the composition of the whole organization in terms of tenure is considered. The time a person has worked on a research team tells us less about the performance of the team than does the average time team members have worked on the particular project together (1985, p. 79).

Pfeffer thus argues that organizations can never be fully understood simply as aggregations of individuals. Social aggregates take on meaning of their own, and have influences of their own: the whole is more than the sum of its parts. This point is critical to the work we have reviewed here, and is the major premise behind the conceptual agenda we propose in this chapter.

Demographic research in organizations is distinctively attractive because it is both empirically feasible *and* resonant of fundamental social processes sometimes considered out of the reach of verifiable research.[19] Unit-level demographic factors, such as seniority distributions, are easily measurable constructs and therefore hold clear appeal for empirical researchers. What is more, demographic factors "stand in for," imply, or affect a variety of less visible, less measurable individual and interpersonal psychological processes and constructs. For example, because cohorts often share early experiences and contexts, chronological age can reflect social, political, and cultural values and attitudes, as well as individuals' ways of thinking (Mannheim, 1952; Ryder, 1965).[20] Clearly, our analyses should reflect the undeniable potential importance of demography in higher education.

Of course, how policy makers feel about and respond to demographic influences can vary. We have not ventured into the normative aspects of faculty seniority distributions here. Others have. Clark Kerr (1994, pp 139-140) argues that among the "rules of conduct appropriate to the effective advancement of knowledge, and to the integrity of teachers in relations with students and of scholars in their relations with other scholars" is "the full acceptance of the obligation, within departments, to seek a reasonable balance of colleagues by age...". In their much-cited book on the U.S. professoriate, Bowen and Schuster (1986) suggest that balanced faculty seniority distributions are often a good thing. Institutions, they note, can and should plan and manage their faculty personnel policies toward a balanced age distribution of their faculties by facilitating timely retirement and by recruiting people selectively by age groups. In saying this, these authors are arguing against institutions' passive acceptance of whatever fate labor markets and demographic

[19]Many analysts doubt the validity of respondents' reports of subjective mental states like attitudes and aspirations, preferring "objective" constructs like behavioral and demographic characteristics (e.g., see Manski, 1993).

[20]Of course, the social effects of age may be more connected with peoples' perceptions or beliefs concerning age than with actual age (Lawrence, 1987, 1988). Like gender (see Ely, 1995), age is a socially constructed characteristic as well as a biological characteristic.

conditions might deal them.[21] We certainly see room for organizational action in this arena, but are cautious about prescribing one direction for demographically sensitive actions among academic institutions. In some settings at some times, heterogeneity in faculty seniority may not foster the most desirable organizational outcomes.[22] There are probably benefits and costs inherent in almost any demographic composition (short of contexts of outright discrimination). Policy choices should be based in the specific realities of different organizational contexts. For that reason, we avoid policy prescriptions here.

We heartily endorse an *analytic* prescription, however: regardless of whether it is from a normative or a more neutral perspective, seniority distributions deserve more attention in our field. Organizational analyses in higher education often overlook demographic factors, and demographic analyses in higher education often overlook organizational factors. Perhaps this is because demographic change, although constant, tends to occur incrementally and to be felt deeply in an organization only over relatively lengthy time periods. Demographic change lacks the drama of, say, sudden retrenchment. Whatever the reason, demographic change can powerfully affect structure, processes, and relationships within academic organizations, and we believe timely analysis of these effects is imperative.[23]

There is no denying that both the professoriate and the academic organizations that employ them are in a period of dramatic demographic transition. On many campuses, the last of one prominent generation of faculty is leaving and a new generation is arriving. What opportunities do these changes bring for higher education, and what problems do they raise? For example, does the infusion of younger, junior faculty into a department with a dominant, established cohort of senior faculty lead to a greater need to formalize and standardize rules and procedures? Does heterogeneity in seniority among department faculty lead to more strains in decision making? Does numerical dominance of one cohort, whether old or young, lead to the acceptance of one, unchallenged "enacted"[24] sense of the department's mission, strategies, external context, and the like? Does numerical dominance of one cohort lead to greater turnover

---

[21]Those interested in organizational re-engineering often argue that the infusion of younger colleagues is essential to success in planning efforts. For example, in restructuring efforts in administrative units at M.I.T., the involvement of staff at least three levels below the unit head is organizationally mandated (Lydia S. Snover, Senior Planning Officer for Institutional Research at M.I.T., personal communication, May 8, 1996).

[22]For some possible negatives of heterogeneity in organizations, see Bettenhausen's literature review (1991).

[23]We are currently beginning our own research program in this area and certainly hope to contribute to future advances in that way, but this review presents a far more comprehensive agenda than we could ever hope to address in our own empirical work. By writing in this form, we are hoping to encourage professional interest and ongoing dialogue regarding a topic we feel will be emerging as increasingly central for postsecondary education.

[24]The term is Karl Weick's (1977). The implication here is that effectiveness might be diminished by a lack of demographic balance.

among members of other cohorts? More broadly, to what extent do the propositions of generic organization theory apply to academe?

To address these largely unresolved questions, this review has illustrated the current contours of this field of inquiry and suggested some specific directions for future theory development and research. Ideally, the final product of pursuing this research direction will be greater understanding of the ways in which the increasingly familiar organizational theme of demographic change is played out within the special context of higher education.

References

Alpert, D. (1985). Performance and paralysis: The organizational context of the American research university. *Journal of Higher Education* 56: 241-281.

Ancona, D. G., and Caldwell, D. F. (1992). Demography and design: Predictors of new product team performance. *Organization Science* 3 (3): 321-341.

Anderson, M. (1990). Resource dependencies of academic departments: A demographic perspective. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Bacharach, S., and Bamberger, P. (1992). Alternative approaches to the examination of demography in organizations. In P. Tolbert and S. B. Bacharach (eds.), *Research in the Sociology of Organizations*, Volume 10. Greenwich, CT: JAI Press.

Bacharach, S., and Tolbert, P. (1992). Introduction to demography volume. In P. Tolbert and S. B Bacharach (eds.), *Research in the Sociology of Organizations*, Volume 10. Greenwich, CT: JAI Press.

Baird, L. (1986). What characterizes a productive research department? *Research in Higher Education* 25: 211-225.

Baird, L. (1990). The melancholy of anatomy: The personal and professional development of graduate and professional school students. In J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*, Vol. VI. New York: Agathon Press.

Baldridge, J. V. (1971). *Power and Conflict in the University.* New York: John Wiley.

Baldwin, R.G. (1979). Adult and career development: What are the implications for faculty? *Current Issues in Higher Education,* Number 2. Washington, DC: American Association for Higher Education.

Bantel, K. A. (1994). Strategic planning openness: The role of top team demography. *Group and Organization Management* 19 (4): 406-424.

Baron, J., Davis-Blake, A., and Bielby, W. (1986). The structure of opportunity: How promotion ladders vary within and among organizations. *Administrative Science Quarterly* 31: 248-273.

Bayer, A. E. (1973). *Teaching Faculty in Academe: 1972-73.* A.C.E. Research Reports, Vol. 8, No. 2. Washington, DC: The American Council on Education.

Becher, T. (1989). *Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines.* Bristol, Pennsylvania and London: The Society for Research into Higher Education and Open University Press.

Berger, M. C. (1984). Cohort size and the earnings growth of young workers. *Industrial and Labor Relations Review*, 37 (4): 582-591.

Bettenhausen, K. L. (1991). Five years of groups research: What we have learned and what needs to be addressed. *Journal of Management* 17 (2): 345-381.

Biglan, A. (1973). Relationships between subject matter characteristics and the structure and output of university departments. *Journal of Applied Psychology* 57 (3): 204-213.

Binder, Barbara J. (1995). Science policy in a knowledge-based economy—New challenges to academic and federal traditions in Switzerland. Invited address to the Euro-

pean Association for Institutional Research, Zurich, August 27, 1995.
Blau, P. M. (1973). *The Organization of Academic Work.* New York: John Wiley.
Blau, P. M. (1977). *Inequality and Heterogeneity.* New York: The Free Press.
Bowen, H. R. (1980). *The Costs of Higher Education.* San Francisco: Jossey-Bass.
Bowen, H. R, and Schuster, J. (1986). *American Professors: A National Resource Imperiled.* New York: Oxford University Press.
Bowen, W. G., and Rudenstine, N. L. (1992). *In Pursuit of the Ph.D.* Princeton, New Jersey: Princeton University Press.
Boyer, C., and Lewis, D. R. (1985). *And on the Seventh Day: Faculty Consulting and Supplemental Income.* (ASHE-ERIC Higher Education Reports, no. 85-3). Washington D.C.: George Washington University.
Braxton, J. M. (1986). Social control in the academic profession. In John C. Smart, ed., *Higher Education: Handbook of Theory and Research,* Vol. II. New York: Agathon Press.
Braxton, J. M., and Hargens, L. L. (1996). Variation among academic disciplines: Analytical frameworks and research. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research, Volume 11* (pp. 1-46). New York: Agathon.
Breneman, D., and Youn, T. (eds.). (1988). *Academic Labor Markets and Careers.* New York: The Falmer Press.
Bryk, A. S., and Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Newbury Park: Sage Publications.
Burke, D. L. (1986). Change in the academic market-place: A study of faculty mobility in the 1980's. Unpublished doctoral dissertation, University of North Carolina, Chapel Hill.
Burke, D. L. (1991). Faculty mobility in an organizational context. In J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research,* Vol. VII. New York: Agathon Press.
Carnegie Foundation for the Advancement of Teaching, The. (1990). Trendlines: Early faculty retirees: Who, why and with what impact? *Change* 22: 31-34.
Chronicle of Higher Education, The. (1994, September 1). *Chronicle of Higher Education Almanac* 38: 1. Washington, DC: Author.
Clark, B. R. (1983). *The Higher Education System.* Berkeley: University of California Press.
Clark, B. R. (1987). *The Academic Life: Small Worlds, Different Worlds.* New Jersey: The Carnegie Foundation for the Advancement of Teaching.
Clark, S. M., and Corcoran, M. (1987, Winter). The professoriate: A demographic profile. *Phi Delta Kappan* LXVII: 28-32.
Clark, S. M., and Lewis, D. R. (eds.). (1985). *Faculty Vitality and Institutional Productivity: Critical Perspectives for Higher Education.* New York: Teachers College Press.
Clark, S. M., and Lewis, D. R. (1988). Faculty vitality: Context, concerns, and prospects. In J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research* Vol. IV. New York: Agathon Press.
Cohen, M. D., and March, J. G. (1986). *Leadership and Ambiguity.* Second Edition. Boston: Harvard Business School.
Coleman, J. S. (1973). The university and society's new demands upon it. In Carl Kaysen (ed.), *Content and Context: Essays on College Education.* New York: McGraw-Hill.
Coombs, G. (1992). Organizational demography: Implications for the organizational development practitioner. In P. Tolbert and S. B Bacharach (eds.), *Research in the Sociology of Organizations* Volume 10. Greenwich, CT: JAI Press.
Corwin, R. G. (1969). Patterns of organizational conflict. *Administrative Science Quarterly* 14: 507-520.
Davis, K. (1940). The sociology of parent-youth conflict. *American Sociological Review*

5: 523-535.

Davis-Blake, A. (1992). The consequences of organizational demography: Beyond social integration effects. In P. Tolbert and S. B Bacharach (eds.), *Research in the Sociology of Organizations* Volume 10. Greenwich, CT: JAI Press.

Deutsch, M. (1973). *The Resolution of Conflict: Constructive and Destructive Processes*. New Haven: Yale University Press.

Dressel, P., Johnson, F., and Marcus, P. (1970). *The Confidence Crisis: An Analysis of University Departments*. San Francisco: Jossey-Bass.

Ely, R. (1995). The power of demography: Women's social constructions of gender identity at work. *Academy of Management Journal* 38: 589-634.

Feldman, D. (1994). The decision to retire early: A review and conceptualization. *Academy of Management Review* 19: 285-311.

Ferris, G.R., Judge, T.A., Chachere, J.G., and Liden, R.C. (1991). The age context of performance-evaluation decisions. *Psychology and Aging* 6 (4): 616-622.

Finkelstein, M. (1984). *The American Academic Profession:.A Synthesis of Social Scientific Inquiry Since WW II*. Columbus, OH: Ohio State University Press.

Finkelstein, S., and Hambrick, D.C. (1990). Top-management-team tenure and organizational outcomes: The moderating role of managerial discretion. *Administrative Science Quarterly* 35: 484-503.

Foner, A. (1988). Age inequalities. In M. White Riley (ed.), *Social Structures and Human Lives*. California: Sage Publication Co.

Goodman, P. (1962). *The Community of Scholars*. New York: Random House.

Gouldner, A. W. (1957). Cosmopolitans and locals: Towards an analysis of latent social roles, Part 1. *Administrative Science Quarterly* 1: 281-306.

Gouldner, A. W. (1958). Cosmopolitans and locals: Towards an analysis of latent social roles, Part 2. *Administrative Science Quarterly* 2: 444-480.

Gusfield, J. R. (1957). The problem of generations in an organizational structure. *Social Forces* 35: 323-330.

Hacker, A. (1992, March 4). Too many full professors: A top-heavy pyramid. *The Chronicle of Higher Education* 38 (26): pp. B1-B2.

Hage, J. (1980). *Theories of Organizations*. New York: John Wiley and Sons, Inc.

Hage, J. and Aiken, M. (1970). *Social Change in Complex Organizations*. New York: Random House.

Hammond, P., and Morgan, H. (eds.). (1991). *Ending Mandatory Retirement for Tenured Faculty: The Consequences for Higher Education*. A report of the National Research Council. Washington, DC: National Academy Press.

Haveman, H. A. (1995). The demographic metabolism of organizations: Industry dynamics, turnover, and tenure distributions. *Administrative Science Quarterly* 40 (4): 586-618.

Hearn, J. C., Anderson, M. S., and Eck, J. C. (1996). Describing academic units: Alternative indicators of faculty composition. Paper presented at the annual joint meeting of the Southern Association for Institutional Research and the Society for College and University Planning, Mobile, Alabama, October 1996.

Hearn, J. C. and Heydinger, R. B. (1985). Scanning the external environment of a university: Objectives, constraints, and possibilities. *Journal of Higher Education* 56: 419-445.

Hellweg, S. A., and Churchman, D. A. (1981). The academic tenure system: Unplanned obsolescence in an era of retrenchment. *Planning for Higher Education* 10: 16-18.

Jackson, S. E., Brett, J. F., Sessa, V. I., Cooper, D. M., Julin, J. A., and Peyronnin, K. (1991). Some differences make a difference: Individual dissimilarity and group heterogeneity as correlates of recruitment, promotions, and turnover. *Journal of Applied Psychology* 76: 675-689.

Kanter, Rosabeth Moss. (1977). *Men and Women of the Corporation*. New York: Basic

Books.

Katz, Ralph. (1982). The effects of group longevity on project communication and performance. *Administrative Science Quarterly* 27: 81-104.

Kerr, Clark. (1972). *The Uses of the University.* New York: Harper and Row.

Kerr, Clark. (1994). *Higher Education Cannot Escape History: Issues for the Twenty-first Century.* Albany, NY: State University of New York Press.

Kinney, D., and Smith, S. (1992). Age and teaching performance. *Journal of Higher Education* 63: 282-302.

Kogan, M. (1984). The political view. In Burton R. Clark (ed.), *Perspectives on Higher Education: Eight Disciplinary and Comparative Views.* Berkeley: University of California Press.

Kosnik, R. (1990). Effects of board demography and directors' incentives on corporate greenmail decisions. *Academy of Management Journal* 33: 129-150.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions.* Second Edition. Chicago: University of Chicago Press.

Lawrence, B. (1987). An organizational theory of age effects. In N. DiTomaso, and S. Bacharach (eds.), *Research in the Sociology of Organizations*, Volume 5. Greenwich, CT: JAI Press.

Lawrence, B. (1988). New wrinkles in the theory of age: Demography, norms, and performance ratings. *Academy of Management Journal* 31 (2): 309-337.

Lawrence, J., and Blackburn, R. (1988). Age as a predictor of faculty productivity. *Journal of Higher Education* 59 (1): 22-38.

Lee, B. (1983). Faculty trends and projected needs. In B. Baldwin and R. Blackburn (eds.), *College Faculty: Versatile Human Resources in a Period of Constraint.* San Francisco: Jossey-Bass Publishing Co.

Levin, S., and Stephan, P. (1989). Age and research productivity of academic scientists. *Research in Higher Education* 30 (5): 531-549.

Lewis, D. R., and Becker, W. E. (eds.). (1979). *Academic Rewards in Higher Education.* Cambridge: Ballinger.

Lodahl, J. B., and Gordon, G. (1972). The structure of scientific fields and the functioning of university graduate departments. *American Sociological Review* 37: 57-72.

Louis, K. S., Blumenthal, D., Gluck, M. E., and Stoto, M. A. (1989). Entrepreneurs in academe: An exploration of behaviors among life scientists. *Administrative Science Quarterly* 34: 110-131.

Lozier, G. G., and Dooris, M. J. (1987). Is higher education confronting faculty shortages? Paper presented at the Annual Meeting of the Association for the Study of Higher Education, Baltimore, November, 1987.

Malos, S. B. and Campion, M. A. (1995). An options-based model of career mobility in professional service firms. *Academy of Management Review* 20 (3): 611-644.

Mannheim, K. (1952). *Essays on the Sociology of Knowledge.* London: Oxford University Press.

Manski, C. F. (1993). Adolescent econometricians: How do youth infer the returns to schooling? In C. Clotfelter and M. Rothschild (eds.), *Studies of Supply and Demand in Higher Education.* Chicago: University of Chicago Press.

March, J. G., and Olsen, J. P. (1979). *Ambiguity and Choice in Organizations.* Bergen, Norway: Universitetsforlaget.

March, J. G., and Simon, H. A. (1958). *Organizations.* New York: John Wiley and Sons, Inc.

Massy, W., Wilger, A., and Colbeck, C. (1994, July/August). Department cultures and teaching quality: Overcoming hollowed collegiality. *Change*: 11-20.

McCain, B., O'Reilly, C., and Pfeffer, J. (1983). The effects of departmental demography on turnover: The case of the university. *Academy of Management Journal* 26: 626-641.

McNeil, K., and Thompson, J. D. (1971). The regeneration of social organizations. *American Sociological Review* 36: 624-637.

Merton, R. G. (1968). The Matthew effect in science. *Science* 159: 56-63.

Messeri, P. (1988). Age, theory choice, and the complexity of social structure. In M. White Riley (ed.), *Social Structure and Human Lives*. California: Sage Publication Co.

Millett, John. (1962). *The Academic Community*. New York: McGraw-Hill.

Milliken, F. J., and Martins, L. L. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Academy of Management Review* 21 (2): 402-433.

Mittman, B. (1992). Theoretical and methodological issues in the study of organizational demography and demographic change. In P. Tolbert and S. B Bacharach (eds.), *Research in the Sociology of Organizations*, Volume 10. Greenwich, CT: JAI Press.

O'Reilly, C. A. III, Caldwell, D. F., and Barnett, W. P. (1989). Work group demography, social integration, and turnover. *Administrative Science Quarterly* 34: 21-37.

Peterson, M. W. (1976). The academic department: Perspectives from theory and research. In J. C. Smart and J. R. Montgomery (eds.), *Examining Department Management, New Directions for Institutional Research*, Number 10. San Francisco: Jossey-Bass.

Pfeffer, J. (1981a). *Power in Organizations*. Marshfield, Massachusetts: Pitman.

Pfeffer, J. (1981b). Some consequences of organizational demography: Potential impacts of an aging work force on formal organizations. In S. Kiesler, J. Morgan, and V. Oppenheimer (eds.), *Aging: Social Change*. New York: Academic Press.

Pfeffer, J. (1983). Organizational demography. In L. Cummings and B. Staw (eds.), *Research in Organizational Behavior*, Volume 5. Greenwich, Connecticut: JAI Press.

Pfeffer, J. (1985). Organizational demography: Implications for management. *California Management Review* 28: 67-81.

Pfeffer, J., and Davis-Blake, A. (1987). The effect of the proportion of women on salaries: The case of college administrators. *Administrative Science Quarterly* 32: 1-24.

Pfeffer, J. and Langton, N. (1988). Wage inequality and the organization of work: The case of academic departments. *Administrative Science Quarterly* 33: 588-606.

Pfeffer, J., and Moore, W. (1980). Average tenure of academic department heads: The effects of paradigm, size, and departmental demography. *Administrative Science Quarterly* 25: 387-406.

Pfeffer, J., and O'Reilly, C. A. (1987). Hospital demography and turnover among nurses. *Industrial Relations* 26: 158-173.

Ramsey, V., and Dodge, L. (1983). Some empirical issues in research on academic departments: Homogeneity, aggregation, and levels of analysis. *Research in Higher Education* 18: 409-419.

Reed, T. L. (1978). Organizational change in the American foreign service, 1925-1965: The utility of cohort analysis. *American Sociological Review* 43: 404-421.

Reskin, B. F. (1979). Review of the literature on the relationship between age and scientific productivity. In Commission on Human Resources, National Research Council (ed.), *Research Excellence Through the Year 2000*. Washington, DC: National Academy of Sciences.

Riley, M., Foner, A., and Waring, J. (1988). Sociology of age. In N. Smelser (ed.), *Handbook of Sociology*. California: Sage Publication Co.

Ryder, N. B. (1965). The cohort as a concept in the study of social change. *American Sociological Review* 30: 843-861.

Smart, J. C. (1990). A causal model of faculty turnover intentions. *Research in Higher Education* 31: 405-424.

Smart, J. C. and Elton, C. F. (1975). Goal orientations of academic departments: A test of Biglan's model. *Journal of Applied Psychology* 60 (5): 580-588.

Smith, K., Smith, K., Olian, J., Sims, H., O'Bannon, D., and Sully, J. (1994). Top management team demography and process: The role of social integration and communication. *Administrative Science Quarterly* 39: 412-438.

Stadtman, V. A. (1980). *Academic Adaptations*. San Francisco: Jossey-Bass.

Stewman, S. (1988). Organizational demography. In W. Scott, and J. Blake (eds.), *Annual Review of Sociology* 14. Palo Alto, CA: Annual Reviews, Inc.

Tierney, W. G., and Rhoads, R. A. (1993). *Enhancing Promotion, Tenure and Beyond: Faculty Socialization as a Cultural Process*. (ASHE-ERIC Higher Education Reports, no. 93-6). Washington, D.C.: George Washington University.

Tolbert, P., and Oberfield, A. (1991). Sources of organizational demography: Faculty sex ratios in colleges and universities. *Sociology of Education* 64: 305-315.

Tolbert, P., Simons, T., Andrews, A., and Rhee, J. (1995). The effects of gender composition in academic departments on faculty turnover. *Industrial and Labor Relations Review* 48 (3): 562-579.

Tsui, A., and Ashford, S. (1991). Reactions to demographic diversity: similarity-attraction or self-regulation. Paper presented at the annual meeting of the Academy of Management Meeting, Miami, 1991.

Tsui, A., and O'Reilly, C. (1989). Beyond simple demographic effects: The importance of relational demography in superior-subordinate dyads. *Academy of Management Journal* 32 (2): 402-423.

Tuma, N. B., and Hannan, M. T. (1984). *Social Dynamics: Models and Methods*. Orlando, FL: Academic Press.

U. S. Department of Health, Education, and Welfare. (1965). *Digest of Education Statistics–1995*. Division of Educational Statistics and Bureau of Educational Research and Development. Washington, DC: U.S. Government Printing Office.

U. S. Department of Education. (1981). *Digest of Education Statistics—1981*. National Center for Education Statistics, Office of Educational Research and Improvement. Washington, DC: U.S. Government Printing Office.

U. S. Department of Education. (1993). *Digest of Education Statistics—1993*. National Center for Education Statistics, Department of Educational Research and Improvement. Washington, DC: U.S. Government Printing Office.

U. S. Department of Education. (1995). *Digest of Education Statistics—1995*. National Center for Education Statistics, Department of Educational Research and Improvement. Washington, DC: U.S. Government Printing Office.

Van Dyne, L., and Jehn, K. (1992). The effects of diversity on group conflict and group communication. Paper presented at the annual meeting of the International Association for Conflict Management, Minneapolis, MN, 1992.

Wagner, W., Pfeffer, J., and O'Reilly, C. (1984). Organizational demography and turnover in top-management groups. *Administrative Science Quarterly* 29: 74-92.

Wallace, W. (1988). Toward a disciplinary matrix in sociology. In N. Smelser (ed.), *Handbook of Sociology*. California: Sage.

Waring, J. (1975). Social replenishment and social change. *American Behavioral Scientist* 19: 237-256.

Weick, K. E. (1977). Repunctuating the problem. In P. S. Goodman, J. M. Pennings, and Associates (eds.), *New Perspectives on Organizational Effectiveness*. San Francisco: Jossey-Bass.

Westphal, J. D., and Zajac, E. J. (1995). Who shall govern?: CEO/board power, demographic similarity, and new director selection. *Administrative Science Quarterly* 40 (1): 60-83.

Whetten, D., and Bettenhausen, K. (1987). Diversity in university governance: Attitudes, structure, and satisfaction. In J.C. Smart (ed.), *Higher Education: Handbook of Theory and Research*, Vol. III. New York: Agathon Press.

Wiersema, M., and Bantel, K. (1992). Top management team demography and corporate strategic change. *Academy of Management Journal* 35: 91-121.

Wiersema, M., and Bird, A. (1993). Organizational demography in Japanese firms: Group heterogeneity, individual dissimilarity, and top management team turnover. *Academy of Management Journal* 36: 996-1025.

Willett, J. B., and Singer, J. D. (1991). From whether to when: New methods for studying student dropout and teacher attrition. *Review of Education Research* 61: 407-450.

Zajac, E. J., Golden, B. R., and Shortell, S. M. (1991). New organizational forms for enhancing innovation: The case of internal corporate joint ventures. *Management Science* 37: 170-184.

Zajac, E. J., and Westphal, J. D. (1996). Who shall succeed?: How CEO/board preferences and power affect the choice of new CEOs. *Academy of Management Journal* 39 (1): 64-90.

Zenger, T. R., and Lawrence, B. S. (1989). Organizational demography: The differential effects of age and tenure distributions on technical communication. *Academy of Management Journal* 32: 353-376.

Feminist Teaching in Higher Education

Becky Ropers-Huilman[1]
Louisiana State University

While teaching has long been an integral part of postsecondary institutions in this country, it has not always been stable or consistent in form or function. Rather, teaching in higher education has been continually challenged by questions that have had the potential to dramatically alter its shape. What knowledge is important to produce, embrace and disseminate? What are the proper roles of teachers and students within the contexts of universities and colleges? What methods are associated with desired outcomes? What priority is teaching for faculty members in a variety of higher education settings?

Within the last several decades a new approach to university teaching has emerged. Feminist teaching has been defined and redefined by the increasing numbers of people who have participated in women's movements and have disrupted and strengthened their work in colleges and universities through feminist politics.[2] In this chapter, I will examine feminist teaching to describe teaching practices of feminists in postsecondary institutions and some of the appendant problems. These practices share an attention to power relations in society and classrooms and, further, consist of efforts to explore the ramifications of those power relations. Further, feminist teaching implies a political commitment to work toward more equitable and responsive learning and teaching environments. Throughout, I emphasize that while feminist teachers regularly consider common issues, they also choose very different approaches to teaching and learning.

I wrote this chapter with the following intentions. I want to focus on the diverse perspectives of feminist scholars because I believe that feminists who participate in higher education contexts have both a great deal to contribute to and learn from their non-feminist colleagues. Through this piece, I hope to encourage and facilitate those interactions. Feminist education literature encompasses and supports richly

---

[1]I would like to thank Clif Conrad, Wendy Kohli, Petra Munro, Philip Bennett, and Stefanie Costner for their close readings of and helpful comments on this work.

[2]Feminist work outside academic circles has undoubtedly shaped feminist teaching practices as well. Additionally, feminist education occurs in a wide variety of settings outside of the academy. However, within this context, I will focus most closely on how feminist teaching has developed within academic environments.

J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*

diverse strands of thought; conversations and concerns found in such literature can teach both feminist educators and other educational participants about the implications of and possibilities for various educational practices and beliefs. By focusing closely on feminist teaching itself, I seek to encourage those identifying with other pedagogical strategies to examine and learn from the similarities and differences between their positions and those I present here.

This chapter is further grounded in the assumption that teaching is a political act. Feminist teachers, as other actors in political spheres, use strategies that were conceived through interactions with a variety of sources. As they shape their work, they strive to be consciously aware of multiple and conflicting political purposes of their practices. In this piece, I emphasize that, like feminism, feminist teaching practices are not unitary, nor are they without complexity. Instead, feminist educators continue to experience and produce intense discussions about the ramifications of their diverse participation in educational settings.

In an effort to embrace these intentions, I have outlined this chapter in the following way. First, I generally discuss feminist perspectives on teaching, asserting that some commonalities exist among various strands of feminism, and that those common threads challenge traditional ways of teaching and learning. I complicate this assertion, though, by acknowledging the variety of theories and strategies encompassed by feminism. Second, I suggest that feminist perspectives have posed critical questions about teaching and learning in higher education environments. I then discuss each of these critical questions in detail, posing multiple options and ways of conceptualizing teaching. Finally, I discuss how this work has been located in a continuing conversation among feminist educators and other higher education scholars and suggest future directions that this conversation may take.

A NOTE ON METHOD

As with any analysis, it is useful to explore the methods chosen to approach inquiry. In this work, I have been guided both by feminist and poststructural approaches. Feminist thought asserts that knowledge is socially constructed and uses various methods to understand those constructions and their effects. In my analysis for this section, I seek to use the tools offered by deconstruction. While deconstruction has commonly been thought of as developing from the work of Jacques Derrida (Lechte, 1994), it has been adopted and modified to suit various purposes. As Marilyn Frye (1992) explains,

> To deconstruct a concept is to analyze it in a way which reveals its construction—both in the temporal sense of its birth and development over time and in a certain cultural and political matrix, and in the sense of its own present structure, its meaning, and its relation to other concepts (p. 163).

Deconstruction looks at binary oppositions (such as man/woman, teacher/student, oppressed/oppressor), suggesting that such positioning of related terms and perspectives diminishes our abilities to see alternatives to the framework provided (Collins, 1996; orig. 1986). Deconstruction also implies that the positioning of such terms as opposite to or essentially not each other has been done within certain discourses[3] to serve certain purposes. Therefore, power is utilized and implicated in this process (Collins, 1996; orig. 1986; Lather, 1991). As such, "It is crucial that we see how the terms interrelate, how they have been historically constructed as opposites, and how they have been used to justify and naturalize power relations" (Orner, 1992, p. 78).

Related to teaching practices in higher education, a deconstruction of commonly understood concepts may lead teachers to envision further opportunities and possibilities for their interactions with others. When those espousing feminist teaching practices have considered such a deconstruction, critical concepts quickly suggest questions that are not easily answerable. Especially in relation to knowledge, difference and power, roles and purposes are unclear and, consequently, are open for reconceptualization. Through this questioning of critical and difficult concepts feminist teachers are struggling to craft and support alternative perspectives and principles of educational practice in higher education.

FEMINIST THEORIES

"*To think of feminism in the singular is sociologically inappropriate*" (Lazreg, 1990, p. 342).

"*Feminism is the political theory and practice that struggles to free all women: women of color, working-class women, poor women, disabled women, lesbians, old women—as well as white, economically privileged, heterosexual women. Anything less than this vision of total freedom is not feminism, but merely female self-aggrandizement*" (Smith, 1990a, p. 25).

Feminist thought originated from many sources and has taken many forms. Feminism generally considers participants' lived experiences as central to their thinking and, therefore, recognizes that feminist theories and practices may be established and used for many different purposes and viewed from many different perspectives. Feminist theory, then, is generally considered useful to the extent that it is applicable, either directly or indirectly, to individual and lived experience. As Cheshire Calhoun (1995) proposes, "Feminist theorizing no longer makes the essentializing assumptions that 'woman' signifies a set of uni-

---

[3] Discourses are "historically contingent, dynamic, and conflict ridden," consist of Adynamic and productive systems" that "produce knowledge, power, and experts," and include "rules that enable members to identify some statements as true or false" (Flax, 1993, p. 39). In the words of Peter McLaren and Colin Lankshear (1993): "Discourses organize a way of thinking into a way of doing. Unlike language, they have both a subject and an object, and actively shape the social practices of which they are mutually constitutive" (p. 381).

versal commonalities, that all women share a common oppression, and, thus, that a single feminist agenda will equally address all women's needs" (p. 9). And Marilyn Frye (1983) believes, "Any theorist would be a fool to think she could tell another woman exactly how the particularities of that other woman's life reflect, or to what extent they do not reflect, the patterns the theorist has discerned" (p. xiii). Finally, bell hooks (1984) asserts that feminist thought must retain its attention to the uniqueness of individual lives. In her words:

> A central tenet of modern feminist thought has been the assertion that "all women are oppressed." This assertion implies that women share a common lot, that factors like class, race, religion, sexual preference, etc. do not create a diversity of experience that determines the extent to which sexism will be an oppressive force in the lives of individual women.

bell hooks further maintains that, "Sexism as a system of domination is institutionalized but it has never determined in an absolute way the fate of all women in this society" (p. 5). For these reasons, feminist thought would never hope to have all-encompassing perspectives or understandings that served regular and understood purposes for every person who attempted to utilize its offerings.

Given this caveat, though, there are general principles that theorists seem to believe apply to a variety of approaches that feminists currently embrace (Flax, 1996; orig. 1979). For example, many feminisms are concerned with political efforts to address the oppression of women. As Sandra Acker (1987) suggests, "Feminist theories serve a dual purpose, as guides to understanding gender inequality and as guides to action" (p. 421). While the definitions of equality and inequality vary among feminist approaches, each maintains an active political element. Patti Lather (1991) describes, "All feminisms appeal to the powers of agency and subjectivity as necessary components of socially transformative struggle" (p. 28). Feminists are taking an active part in changing the world.

Further, contemporary feminist approaches seek to be inclusive of multiple perspectives in their theorizing and activism. In the words of bell hooks (1984):

> Feminism is the struggle to end sexist oppression. Its aim is not to benefit solely any specific group of women, any particular race or class of women. It does not privilege women over men. It has the power to transform in a meaningful way all our lives (p. 26).

Through embracing and being informed by work focusing on a wide variety of lived experiences, feminism has strengthened its commitment to become a political movement to which the many, rather than the few, can belong.

Many feminists consider part of their activism to be a reconceptualization of the ways children and students are taught to behave as gendered beings in social institutions. This attention to education has generated much discussion on how feminist thought can and does affect teaching. Often, feminism has suggested that educators must reconsider both the knowledge that we use in classrooms, as well as the ways we use that knowledge (Boxer, 1985; Frye, 1992; Gore, 1993; Jipson, Munro, Victor, Froude-Jones, and Freed-Rowland, 1995; Weiler, 1988). A reworking of accepted knowledge and the norms involved in teaching and learning certain knowledge has emerged as a central goal of feminist education. Both content and process are considered in the development of feminist teaching practices.

For each similarity presented here, feminist strands of thought respond with a wide variety of approaches and conceptualizations. Multiple feminist approaches have emerged, in large part, because, "We have repeatedly discovered that we have overlooked or misunderstood the truths of the experience of some groups of women and that we have ourselves been overlooked or misunderstood by some other segment or school of feminist thought" (Frye, 1992, p. 61-62). As such, I highlight the tensions and questions, rather than the resolved "answers," involved in feminist teaching in an effort to allow for wider engagement with and deeper thought about these important concepts.

CRITICAL QUESTIONS: A FEMINIST FRAMING OF ISSUES IN TEACHING

"Tensions can be productive as well as destructive. It is through these very dilemmas, through attempts to solve the unsolvable, through exchanges among feminist frameworks that feminist theory moves ahead" (Acker, 1987, p. 432).

Feminist teaching has grown from different origins and has found differing obstacles and opportunities to overcome and embrace. In this analysis, I consider the growth and development of feminist teaching in higher education institutions. Therefore, I focus in this section on the ways in which many feminist teaching practices and approaches have been woven throughout higher education and its proscribed norms of teaching and learning. Through feminist teaching, feminism and higher education have, in many cases, become intertwined. Those intersections have created vibrant, if sometimes conflicted, patterns of growth.[4]

The issues that I choose to focus on here are those whose outcomes are still contested. However, they are certainly not all-encompassing of those involved in feminist education. Undoubtedly, readers will be able to lengthen the list consid-

---

[4] Many writers have considered feminist and women teachers= struggles in relation to various aspects of academic settings, largely because their classrooms are located within a society that values men differently than women (Barbezat, 1988; Chamberlain, 1991; Hensel, 1991; Minnich, O'Barr and Rosenfeld, 1988; Pounder, 1989; Schuster and Van Dyne, 1985; Simeone, 1987; Uttal, 1990). Other authors have examined issues associated with understanding and reshaping women's roles as students and classroom participants (Belenky et al., 1986; Fiske, 1992; O'Barr and Wyer, 1992; Statham, Richardson, and Cook, 1991). The intense dilemmas presented in these works undoubtedly shape enactments of feminist teaching that are able to occur in formal higher education institutions. They are not, however, the focus of this paper. Rather, I concentrate here most closely on the debates *within* feminist academic communities about the possibilities and challenges for feminist teaching practice.

erably based on their own experiences or those of their colleagues. Neither do I intend to imply that I believe these issues must be resolved. Rather, I argue that these dilemmas, and the options they present, have emerged within a specific historical time period that highlights certain conflicts and norms above others. In other words, these struggles have been named for various purposes. Their preservation may continue to invigorate analyses of politics involved in feminist teaching and in other higher education practices.

I now turn to what I perceive as three of the most perplexing and pervasive points of discussion that are shaping and troubling feminist teaching today, those relating to knowledge, difference and power. Discussions about these concepts are integrally related to many areas of higher education thought, especially those concerning the rationales for our chosen educational strategies. My purpose, therefore, is to assert that those using feminist teaching practices have much to contribute to a broader higher education community that cares very deeply about teaching and learning.

What is the role of knowledge?

The creation and dissemination of knowledge, through research and teaching responsibilities of professors, have long been considered functions of higher education institutions. Debates about how, why and what knowledge is created and disseminated have retained prominent positions in higher education discourses as well. These debates center themselves on various aspects of the question: *Should the creation and dissemination of knowledge be an objective process or should the process focus on creating practical knowledge for activist purposes?*

Higher education scholars have asserted both that knowledge must be objective and unrelated to those engaging with the knowledge (Pelikan, 1992) and that the knower and knowledge are integrally related and should not be separated in intellectual thought processes (Palmer, 1983). They have further suggested that while knowledge in higher education should continue to be constantly evaluated, certain methods of knowledge production and negotiation[5] are more useful than others in the "aspiration to excellence" (Barber, 1992, p. 119). And while some assert higher education's responsibility to the society that supports it, "The ambivalence of universities toward social change persisted—and persists" (Pelikan, 1992, p. 158).

Feminism has furthered this debate in academic settings in a variety of ways. For example, feminists assert that knowledge is political, shifting and always in the process of being constructed. While feminist work generally asserts that feminist scholars should embrace knowledge that could be useful in political contexts, debates about the roles of knowledge and, consequently, the roles of feminist teachers, take on a variety of forms.

---

[5]See Ropers-Huilman (1996a) for a more detailed discussion on the concept of negotiating knowledge in feminist teaching environments.

Knowledge as political

"The feminist teacher who is calling into question accepted beliefs and attitudes about gender no more imposes her own meaning and culture upon the classroom than does the traditional teacher who wants students to accept authority and the status quo" (Weiler, 1988, p. 136-137).

How do feminists enact their politics through their teaching positions in higher education? In this section, I discuss the rationale for an explicitly political agenda that many feminist teachers embrace. Further, I consider how feminist academicians have been urged to enact their politics for and with diverse people both in and outside of academic communities. In doing so, I highlight the challenges and questions brought to bear when one embraces a view of knowledge as political.

Some feminist teachers assert that their politics include a commitment to ensuring a non-discriminatory classroom, respectful language use by all participants, and a use of content that reflects a diversity of perspectives (Ropers-Huilman, in press). Many feminist scholars insist that their work in classrooms is inherently political and that, therefore, teachers should be explicit about their politics (hooks, 1996; orig. 1989; Middleton, 1993; Weiler, 1988). But what form should this politics take and why is it so important to continually examine one's teaching efforts in light of one's political underpinnings and potential effects? Because of the urgency of this question, bell hooks (Vaid et al., 1993) suggests that politics, in teaching and in other feminist practice, should explore links between feminist theory and feminist political strategies. In her words, "I'm interested in how we create a politics of everyday life. That means not only challenging the mass media. It means coming up with some very different strategies: one, for sharing what feminist thinking and practice are, and two, for talking about how you utilize them in your life" (p. 41). Drawing on Freire's work, many other feminist and critical scholars suggest that we must teach through attention to real consequences of and possibilities for our enactments of knowledge (Kenway and Modra, 1992; orig. 1989; Shor, 1993; Weiler, 1988). Knowledge must be political.

The feminist claim that attention to politics in classrooms is important has drawn attention to the difficulty that many scholars have in seeing opportunities for human agency within postmodern feminism.[6] For example, Carol Nicholson (1995) expresses her concern that a postmodern agenda is limiting to the necessary explicitness of political purposes within a feminist classroom as follows:

> It has been charged that postmodernism in education entails neoconservatism in politics. Students who are taught that there is no rational way to justify political change are in effect being told to accept the status quo, an attitude that is ultimately stultifying to

[6] Postmodernism is characterized by a suspension of "metanarratives" that suggest generalizable knowledge for all social experience, and a questioning of the traditional distinctions between subject and object (Lather, 1991; Lechte, 1994).

> thought. Feminists opposed to postmodernism emphasize the need to give students ideals to live for, principles to live by, and a vision of a better future that they can help to achieve (p. 84).

While politics are undoubtedly made more complex by critical and poststructural questionings, they are nevertheless inherent within any educational practice. Choices about knowledge use are made for certain reasons to, hopefully, achieve certain ends or at least bring new questions or perspectives into classroom discourse. Postmodern feminism draws attention to the results and opportunities for several types of feminist educational politics, presenting and exploring the vast and complex terrain that informs those politics.

Another question concerning the politics of knowledge relates to how feminist teachers use their positions and knowledge to serve diverse communities that exist both in and outside of university walls. How can feminist educators be of most use to their various communities? What is their purpose? The debates about academic feminists' presence in the larger feminist community have been substantial. For example, those who practice their feminism in academic settings have been accused of acting in ways that do not address the needs of those in other settings. As bell hooks (1996; orig. 1989) suggests:

> Since the work of feminist theorists necessitates fundamental questioning and critiquing of the ideological structures of the prevailing white-supremacist, patriarchal hegemony, it is fitting that the university be identified as a useful site for radical political work, for feminist movement. It must be remembered that it is not and should not be the only site of such work. Academic women and men engaged in the production of feminist theory must be responsible for setting up ways to disseminate feminist thought that not only transcend the boundaries of the university setting, but that of the printed page as well (p. 57).

Marnia Lazreg (1990) also discusses how feminists in the academy have neglected to uncover some of their own philosophical and developmental bases. In her words: "Knowledge is produced not only within a socioeconomic and political framework but also within an intellectual tradition with stated and unstated assumptions. Although it questions traditional assumptions, academic feminism has often neglected to investigate its own premises" (p. 327). Rather than attending to the immediate concerns of a variety of women, feminist academics have further been accused of using their positions to develop practices that advance theoretical obscurity and serve only those who are similarly positioned (Childers and hooks, 1990; hooks, 1984). Feminist educators, some argue, must insure that their use and development of knowledge serves the communities and ideologies that they intend to foster and support.

Knowledge as personal

"*Ethics benefits from reflection upon our own experience, upon choices we have actually faced*" (Card, 1991, p. 4).

In an effort to create and work with knowledge that can be political as described above, specific experiences outside formal educational environments and knowledge from those experiences are often included in classroom analyses. Many feminists have supported the use of what is commonly understood to be "personal" knowledge in academic teaching and learning settings. This personal knowledge is usually considered to be that which results from individual experience or interactions with others in non-academic settings. Its use often challenges the dichotomy of public and private (or academic and non-academic) in education.

Carmen Luke (1992) is among many other feminist educators who expresses her belief that analysis of or including personal experiences is useful in feminist classrooms. In her words, "It is unquestionably important to give students the analytic tools with which to understand the forces that shape their experience, the first step of which is encouraging students to articulate their experiences and sense of self" (p. 36). Kathleen Weiler (1988) further asserts:

> The realm of common sense is open to critique because of the hegemonic ideology it partially embodies; but common sense itself provides the means of that critique through its own thought processes and practical activities. Thus the contradictions of everyday life and consciousness itself can become the focus of a radical pedagogy (p. 23).

From the perspectives of these educators, personal experience is a necessary analytical piece when considering the effects and implications of what is considered to be "public" knowledge.

As noted earlier, many feminists have drawn on Paulo Freire as a guide for their teaching practices. Freire's work further supports the grounding of education in personal experiences of participants. For example, Freire (1990) believes that, "No pedagogy which is truly liberating can remain distant from the oppressed by treating them as unfortunates and by presenting for their emulation models from among the oppressors. The oppressed must be their own example in the struggle for their redemption" (p. 39). And Ira Shor (1993) explains Freire's insistence on the importance of the interactions between students' own experiences and knowledge bases and those that are newly presented in a formal educational setting. In his words:

> For Freire, teaching and learning are human experiences with profound social consequences. Education is not reducible to a mechanical method of instruction. Learning is not a quantity of information to be memorized or a package of skills to be transferred to students. Classrooms die as intellectual centers when they become delivery systems for lifeless bodies of knowledge. Instead of transferring facts and skills from teacher to students, a Freirean class invites students to think critically about subject matter, doctrines, the learning process, itself, and their society. Freire's social pedagogy defines education as one place where the individual and society are constructed, a social action which can either empower or domesticate students (p. 25).

Feminist teaching shares this attention to participants' lived experiences, suggesting that it is necessary to consider multiple perspectives so that all can teach and learn in personally relevant and meaningful ways.

This is not to say, though, that those using feminist teaching believe women and men, as students in classrooms and in structures, may not experience educational settings and practices differently, just as they experience and interpret intelligence differently. Women and men are often seen as having different relationships with what is understood to be personal and public knowledge. Andrea Dworkin (1992; orig. 1983), for example, insists that women and men are expected to display and engage with certain types of knowledge and experience, based on society's construction of their gender. Further, she believes that intelligence as it is commonly understood is something that women are not expected or, sometimes, even permitted to acknowledge in themselves. In her words:

> A woman must keep her intelligence small and timid to survive. Or she must hide it altogether or hide it through style. Or she must go mad like clockwork to pay for it. She will try to find the nice way to exercise intelligence. But intelligence is not ladylike. Intelligence is full of excesses (p. 105).

Intelligence and knowledge take on different meanings for each individual who engages within them. Personal experience, though, is influenced by societal norms and expectations of persons based on the multiple facets of their identities.

Feminist teaching literature suggests many questions about knowledge for consideration in our struggles as teachers and researchers who seek to improve and understand educational experiences:

- How are power and knowledge constantly interacting in a particular classroom or in this particular institution?
- How can educators better embrace persons from different races and cultures than their own through their acknowledgment and validation of diverse knowledge sources?
- How can feminist teachers explode notions of normalcy embedded in certain practices through their multiple uses and interrogations of knowledge?
- How can teachers examine and understand the political implications of knowledge negotiations that take place in academic environments?

In feminist teaching, interactions with knowledges become circular and dynamic. Personal experience is presented with knowledge from academic settings; each piece of knowledge then challenges the existence of others within this re-forming self. Personal and public together create the meanings that students and teachers alike are tentatively predisposed to embrace. It is through this interchange that political and personal knowledge inspire and interrupt each other, becoming nourishment for continuance.

What is the role of difference?

"*As many feminists have been forced to admit, the recognition of difference and diversity is a political challenge....The achievement of diversity and unity can only result from political effort*" (Banks, Billings, and Tice, 1996; orig. 1993, p. 87).

Difference and equality are two terms that have generated heated debates and grave concerns about everyday practices in higher education. If there is a women's studies program and a women's center, where are the corresponding centers and programs for men? Should some racial groups receive "special" consideration in admissions to colleges and universities? What is happening to the core curriculum that has served our institutions in the past? What does "inclusive curriculum" really mean?

Higher education has been stubbornly resistant to change in the more than 350 years of its development in the United States (Rudolph, 1990; orig. 1962). And while the populations served by these institutions have changed dramatically, the structures that support and characterize them have largely failed to address the corresponding changing needs (Damrosch, 1995). Because "conventional education strives not to locate and understand the self in the world, but to get it out of the way" (Palmer, 1983, p. 35), attention to the differences embodied in current higher education settings may not be immediately addressed or incorporated into what is considered to be "normal" practice in education.

Recently, those involved explicitly in the discourse of equality and difference have largely shared similar themes. Educators generally want to be fair and inclusive. However, those same scholars argue ferociously over the definitions of fairness, equality and difference, and the ways that those dynamic concepts should be addressed in higher education settings (Firing Line, 1996). For example, Dinesh D'Souza (1992) points out the drawbacks of several efforts to include previously excluded groups as participants in higher education settings. He proposes that current efforts to include multiple perspectives, to recognize difference and to treat individuals differently based upon that difference, are actually exclusive, unfair processes that deny certain opportunities to those who are actually "better" in an equal system. As such, "There is no uniform standard of justice which, as Aristotle observed, is the only lasting basis for community" (p. 50). "Preferential treatment" policies, therefore, intensify hostilities between races or other identity groups, and work against communities that higher education institutions are trying to build.

Other scholars suggest that preferential treatment is necessary both because of racially-based biases that have existed in the past—which supported the exclusion of large groups of people—and because our current curriculum and educational experiences are still biased to serve the needs of white, heterosexual, middle-class men (Barber, 1992; Holland and Eisenhart, 1990; B. Smith, 1990). Since white women, lesbians or gay men, working-class persons, and women and men of color have traditionally been, and continue to be, excluded from some parts of educational discourse, additional resources should be provided to ensure a result of relatively fair—or equal—treatment.

Feminist scholars have been active in conversations about equality and difference. As Rosemarie Tong (1989) illustrates, their contributions are complex.

> Some feminists worry that an overemphasis on difference may lead to intellectual and political disintegration. If feminism is to be without any standpoint whatsoever, it becomes difficult to ground claims about what is good for women. It is a major challenge to contemporary feminism to reconcile the pressures for diversity and difference with those for integration and commonality (p. 7).

In feminist teaching and scholarship, discussions about roles of equality and difference have been framed by these questions: *Since feminist arguments have been made using the concepts of both equality and difference, how is it useful or harmful to continue operating as if those concepts were a dichotomous pair? Put differently, do we have to insist that we are all the same in order to be treated equally? What would it mean to feminist educational practice if teachers were to "problematize" the relationship of equality and difference in their practices?*

Many scholars have suggested that women must decide whether they want to be, and be treated as, "equal to" or "the same as" men or "different" from men. As a result, "The sameness/difference framework calls for simple yes or no answers that requires suppressing counter-examples" (Minow, 1990, p. 153). Marianne Hirsch and Evelyn Fox Keller (1990) support the tenaciousness of this debate by suggesting the limitations of its options. In their words, "Too often, the work of exploring differences among commonalities, and commonalities within difference, has been displaced by a defensive and anxious need to 'choose sides'" (p. 384). Yet, in a compelling argument, Joan Wallach Scott (1990) considers the meanings and intents of these limitations and concludes that, "Equality is not the elimination of difference, and difference does not preclude equality" (p. 138).

In this section of the chapter, I discuss the ways that those in feminist education have attempted to circumvent limitations of the equality versus difference debate. I assert that through a heightened attention to the many categories of and possibilities for "difference," they have problematized the notion that sameness is a necessary precursor to equality. As such, I draw on many feminist educators and theorists to consider the changes that might occur when feminist teaching wrestles with the parameters of equality and difference.

Equality in classrooms

Many feminist scholars believe that men and women, as well as people with other different identity characteristics, have not been treated equally within educational settings (Dworkin, 1992; orig. 1983; hooks, 1994; Middleton, 1993; Stitt, 1988). As such, those scholars advocate for attempts to create environments in which all participants can learn from and contribute to the ongoing discourse regardless of their identities (Belenky, Clinchy, Goldberger, and Tarule, 1986; Sadker and Sadker, 1994; Stitt, 1988). These scholars, though, are aware that a quest for equality is not unproblematic. Further, those who are attempting to ensure equality in their classrooms do not expect as a matter of course that all others will embrace their ambitions.

Equality has long been an established value of feminist thought. Yet for many reasons, the concept of equality quickly becomes muddied by the implications of its uses. For example, in a conversation where Gloria Steinem and bell hooks (Vaid et al., 1993) discuss the principles of equal participation in feminist groups, Steinem uses the following words: "We came out of small groups where the whole point was that you listened with respect to other people who spoke from the 'I.' You didn't try to tell somebody else what to do. You didn't try to judge them. You just each spoke the truth....It's respect for individual experience." hooks replies: "That's exactly the model that nauseates me. I hate being in a room where we're made to feel like everybody has an equal voice. Because I've seen that model close down dissent, make it appear that all opinions are equal" (p. 41). These two established and prominent feminists disagreed on the methods through which to best ensure equal participation and respect during interactions. Paths to equality are indeed full of choices and complexity.

Christine Sypnowich (1996; orig. 1993) also questions whether attempts to establish equality generate a false and harmful "universalizing" or "essentializing" of experience. In other words, is it possible to establish a movement of any kind without a central understanding about the common identity of purpose that is shared? In her words:

> From a philosophical point of view, it's not clear whether universalism of some kind is avoidable, in any case. After all, the fate that met feminists, who, having rejected the false universalism of androcentric discourse were then accused of false universalism in the face of race, class, etc., is a fate that can meet the assertions of identity that come after them; difference unleashes an endless cycle of accusations and inclusions. . . . Ultimately there are as many differences as selves, and thus our invocations of difference always risk essentialism, wherein we reify a certain identity and proclaim its immutable nature, without attention to the differences within the identity itself, or the damage done to the new "other" the reclaimed identity leaves in its wake (p. 285).

Can feminist movement establish equality? What are the dangers in grounding a movement based on equity principles? As Ann Snitow (1990) states, "Although women differ fundamentally about the meaning and value of 'woman' we all live partly in, partly out of this identity by social necessity" (p. 13). Yet many others have argued that attempting to assert commonality of experience and belief in efforts to achieve equality tends to re-establish the norms and expectations of the dominant cultures which are involved in those efforts (Anzaldua, 1990; Vaid et al., 1993).

Critics of feminist approaches that focus exclusively on equal participation and opportunity pose questions largely on two fronts. First, Ann Snitow (1990) questions this approach for its lack of usefulness in allowing women to occupy positions which they themselves construct, rather than merely reshaping themselves to prove their equality in relation to male standards. In her words:

> The biggest complaint against a feminist demand of "equality" is that this construction means women must become conceptual men, or rather than to have equal rights they will have to repress their biological difference, to subordinate themselves in still new ways under an unchanged male hegemony (p. 26).

Women, in this framework, will be accepted as equals only if they succeed in matching certain expectations of the pre-existing patriarchal systems and institutions. Rather than questioning the wide-ranging impact of power and patriarchy on women, limited views of the relationships between equality and difference suggest ways in which women can succeed within current power relations and patriarchal structures. At no point are women called on to restructure the systems within which they are hoping to achieve equality.

A second critique of approaches defining equality without considering difference questions the essentialism inherent in its principles. If women and men, whether students or teachers, are conceptualized as existing in discrete, identifiable categories, the wide variety of differences within and between those categories will likely be obscured. A poignant example of this conundrum is demonstrated in bell hooks' questioning, "Since men are not equals in white supremacist, capitalist, patriarchal class structure, which men do women want to be equal to?" (Snitow, 1990, p. 26).

A feminist classroom based on the belief that one need only ensure equal chances for participation for all students will have great difficulty, in my mind, achieving its goals. Such an approach suggests that differences between identities and ideas do not affect participants' experiences in teaching and learning settings. As Carmen Luke (1992) suggests,

> To grant equal classroom time to female students, to democratize the classroom speech situation, and to encourage marginal groups to make public what is personal and private does not alter theoretically or practically those gendered structural divisions upon which liberal capitalism and its knowledge industries are based (p. 37).

Attempts to achieve equality within educational settings are doomed to failure if they ignore the personal and diverse experiences of participants. Equal participation or opportunity in pre-existing and unequal structures without attention to differences is not equality.

Identity differences in classrooms

"As a tool of social control, women have been encouraged to recognize only one area of human difference as legitimate, those differences which exist between women and men. And we have learned to deal across those differences with the urgency of all oppressed subordinates" (Lorde, 1992, p. 407).

As stated earlier, the methods through which to approach and include differences within classroom settings have generated much concern and debate in higher education. Especially in relation to the concept of "woman," many in feminist education have suggested that we need to pay more attention to the different identities of those who participate in our classrooms and conversations (Anzaldua, 1990; Brady and Hernandez, 1993; Nicholson, 1995; Snitow, 1990). In some cases, feminist education, in trying to establish equality between women and men, has been complicit in allowing a continued hierarchy, or inequality, to exist—one which reinforces the interests of white, middle-class, heterosexual women and men above all others (Butler, 1990) and results in a "critically important philosophical mistake" (Ginzberg, 1991, p.128).

Recently, several scholars have pointed out this failure and have insisted that it be addressed. For example, Gloria Anzaldua (1990), in explaining why she had once again taken on the task of compiling a book of writings by women of color, writes the following:

> Racism is a slippery subject, one which evades confrontation, yet one which overshadows every aspect of our lives. And because so few (white) people are directly and honestly talking about it, we in the book have once again had to take on the task. Making others "uncomfortable" in their Racism is one way of "encouraging" them to take a stance against it (Anzaldua, 1990, p. xix).

And Marilyn Frye (1992) asserts that feminism must be fueled by the diversity, rather than the similarity, of women's experiences. In her words:

> It is an unforgettable, irreversible and definitive fact of feminist experience that respect for women's experience/voice/perception/knowledge, our own and others', is the ground and foundation of our emancipation—of both the necessity and the possibility of rewriting, recreating, the world. Thus it is only by a violent dishonesty that we could, or can, fail to give credence to women's voices even when they wildly differ and conflict (p. 63).

Finally, Kathleen Rockhill (1993) believes that educators need to question the ways in which we have ignored the implications of our heterosexist society for our classroom and political practices. As she suggests:

> To address, seriously, the question of women's power, it is essential to open up the ways in which we are implicated in institutionalized heterosexism: the ways in which we live heterosexism, not only in our intimate relationships, but also in public settings, where we are also positioned as the sexualized female, and the ways in which our very identity—our sense of self, our subjectivity—has been shaped by heterosexism, as defined through the prisms of class, race, and culture (p. 350).

A central dilemma that arises in feminist work about difference relates to the "ordering" of characteristics that serve as bases for oppression. Largely due to the influences of poststructural feminism and feminist work being done by those who focus on marginalized experiences, much scholarship has recently suggested that identities are composed by the conflation of individual characteristics and experiences. Further, those characteristics often become indistinguishable in their effects. For example, Audre Lorde (1990) discusses the intersections of her black and lesbian identities in this way:

> Just as racist stereotypes are the problem of the white people who believe them, so also are homophobic stereotypes the problem of the heterosexuals who believe them. In other words, those stereotypes are yours to solve, not mine, and they are a terrible and wasteful barrier to our working together. I am not your enemy. We do not have to become each other's unique experiences and insights in order to share what we have learned through our particular battles for survival as Black women (p. 324).

And Gloria Yamato (1990) describes the intersections of multiple identities as they influence and support institutionalized racism.

> Racism is the systematic, institutionalized mistreatment of one group of people by another based on racial heritage....Racism is supported and reinforced by classism, which is given a foothold and a boost by adultism, which also feeds sexism, which is validated by heterosexism, and so it goes on (p. 22).

Feminist scholarship and action has asserted that in order for feminism to achieve a goal of equality, it must also pay close attention to the differences in identities that shape our educational institutions and our world.

Paying attention to differences takes many forms in higher education classrooms. Because one can never know the nature of complexities within participants in educational settings (Ellsworth, 1992; orig. 1989), attention to differences first and foremost means flexibility in educational practices and a willingness to remain open to seemingly non-traditional or "radical" ways of teaching and learning. These radical teaching methods can include a valuing of complexity and an acknowledgment of uncertainty related to a given content area or teaching strategy. Sometimes, conflict resulting from and exploration or heightening of difference can serve as a catalyst for difficult, yet useful, educational exchanges. Other times, differences can serve to further analytical frameworks and suggest alternative options for politics and understanding. Further, differences can help people to see patterns of experience around which to base political action. Marilyn Frye (1992) believes that differences were, in fact, necessary, rather than detrimental, to the development of feminist movement and analyses of women's experiences. In her words:

> The differences of experience and history are in fact necessary to perception of the patterns. It is precisely in the homogeneity of isolation that one cannot see patterns and one remains unintelligible to oneself. What we discover when we break into connection with other women cannot possibly be uniform women's experience and perception, or we would discover nothing. It is precisely the articulation and differentiation of the experiences formulated in consciousness-raising that gives rise to meaning. Pattern discovery and invention requires encounters with difference, with variety. The generality of pattern is not a generality that defeats or is defeated by variety (p. 66).

Identity differences are educationally valuable and have a great potential for teaching and learning in feminist classrooms.

Idea differences in classrooms

Closely related to what I am calling "identity differences" are "idea differences." These are largely intertwined and have deep implications for feminist practice in educational settings.[7] In this section, then, I consider the ways that feminist educators have attempted to address and be inclusive of different perspectives in higher education settings.

Largely at the urging of feminist scholars to consider personal, experiential knowledge as a valid source of classroom material and the insistence of post-structural scholars to understand limitations of our knowing, many in feminist education have advocated for the elevation of attention to differences of view, or multiple realities, in their classrooms (Bee, 1993; Ellsworth, 1992; orig. 1989; hooks, 1996; orig. 1989). Some assert that feminist teachers must embrace a wide variety of ideas in their curricula and urge students to freely contribute their multiple perspectives to class discussions (hooks, 1996; orig. 1989; Sypnowich, 1996; orig. 1993; Weiler, 1988). Others believe that the very ability to know each other's ideas must be problematized because our interactions as teachers and learners within societal structures are limited by and limiting of our abilities to interact "freely" with each other (Ellsworth, 1992; orig. 1989; Flax, 1993). In other words, the notion that teachers and students can ever know each other in complete and authentic ways is suspect.

In their interactions with each other, these claims suggest several ideologies that increasingly serve to add to the complexity of feminist teaching practices. For example, Jane Flax (1993) believes that while we can never be certain of truth or the extent to which constructions of truth influence our interchanges, ideas are put forth within discourses and are validated or invalidated within specific contexts. The implications of this assertion for feminist teaching are complex. If truth is unstable and knowable only within certain discourses, and if we are all acting in and influenced by varying, multiple, and shifting discourses that act dynamically with each other, then the questions of what ideas or truths are proposed is less crucial than the question of why certain ideas were proposed, or brought into the classroom discourse, and why certain ideas were excluded from the same discourse. While "postmodern feminism encourages sensitivity to differences and to a plurality of interpretations of human experience" (Nicholson, 1995, p. 83), it also suggests that "all differences are not equal nor do they deserve the same political consideration" (Flax, 1993, p. 111).

---

[7] I chose to disconnect identity differences from idea differences in this analysis to ensure that the importance of both concepts is highlighted and emphasized. In so doing, though, I risk suggesting that identity and idea differences are two separate and mutually exclusive categories. This is not my belief. Rather, I assert that one strongly, but not absolutely, implicates the other.

Intense struggles within feminist communities have generated perplexing questions for examination by those in higher education related to difference and equality:

- Are women and men being treated equally (given equal attention and opportunity, both as quantitatively and qualitatively measured) in this classroom or educational institution?
- How are women and men treated differently in educational settings? And how does that affect their opportunities in society?
- Do gender neutral or gender specific policies support equality among women and men?
- How are educational experiences implicitly and explicitly guided by heterosexual norms, practices, and expectations?
- How can teachers ensure positive experiences in their classrooms for those coming from racial minority experiences?
- How does race interact with gender, class, sexuality, and other identity characteristics to shape possibilities for and outcomes of educational experiences?
- How have certain identity and idea differences shaped intellectual, social, and political practices in higher education?

Feminist educators must decide: In our attempts to establish political and personal equality in our classrooms, what kinds of differences must we explore? What are our intentions in doing so? How do we privilege some ideas and identities over others? And what are the implications of or paradoxes in those choices? Feminist scholars are fueled by these concerns regularly as they make decisions about their participation in teaching and learning environments.

What is the role of power?

"Perhaps it is more comfortable to treat difference as variations between fairly homogeneous and unrelated blocks. Then one can deny complicity in constructing and being constructed by the difference of the others" (Flax, 1993, p. 6).

Debates about difference have generally revealed much about the power relations embodied and supported by those involved in the debates. In higher education literature, some scholars consider higher education as a way to gain opportunity, thereby increasing one's personal power (D'Souza, 1992; Pelikan, 1992). Others have perceived education itself to be a discursive tool whereby power relations are either maintained or disrupted (Apple, 1993; Giroux, 1993; Ropers-Huilman, 1996b). Largely through examinations of curricular issues and constructions, power structures within teaching and learning settings have grown to be a point of much discussion in higher education literature (Barber, 1992).

Power and difference are closely intertwined in feminist teaching practices. Who has the power to determine which differences are acknowledged and included in curricula? How are different meanings attended to or negated in educational experience? And how are different identities reinforced as playing particular roles in feminist classrooms? How is power itself constantly redefined through the nature of interactions? Recently, the perplexing dilemmas about power in feminist classrooms have been framed by the questions: *What types of power should and do teachers and students have in certain situations? What are the desired outcomes of those power enactments?*

In this section of the chapter, I consider the various arguments and rationales for feminist teachers' maintaining and giving up certain types of power. Further, I problematize the notion that feminist education should strive to establish or achieve a goal of eliminating power considerations from the classrooms. With Gloria Yamato (1990) and others (Gore, 1992; Weiler, 1988), I assert that, "While in one sense we all have power, we have to look at the fact that, in our society, people are stratified into various classes and some of these classes have more privilege than others" (p. 22), and that we must take those power relations into consideration in our educational practices.

Minimizing power in teaching and learning

Several feminist educators believe that power relations in higher education, because of evaluation and assumed determination of content and process, set teachers firmly as having power over students (see Ropers-Huilman, in press, for examples). As such, some feminist and critical educators are concerned with their own uses of power as they enact it in classroom settings (Rockhill, 1993; Weiler, 1993). As one approach of many that they may choose, some argue for a minimizing of the power differential between teachers and students in higher education classrooms. By allowing students to determine the focus of course work and have a large say in evaluation procedures, feminist teachers are drawing attention to the disparity of power relations in classrooms and are seeking to minimize its effects. As Mary Bricker-Jenkins and Nancy Hooyman (1986) write: "A feminist approach goes beyond content and analysis to deal with issues of authority, power, and control and includes a commitment to nonhierarchical and nonoppressive relationships in the learning setting" (p. 41).

In specific situations, generally ones that are closely linked to feminist teaching practices, this power differential takes on additional complexities. For example, several educators believe that when they ask students to work in non-traditional ways, expressing feelings and thoughts about course material, they must also examine their own position in doing so (Middleton, 1995; Orner, 1992; Rockhill, 1993). In other words, educators must examine their reasons for encouraging whatever strategies they choose and look closely at their chosen participation in the resulting interactions. Further, as bell hooks (1994) asserts, faculty are sometimes uncomfortable with cultural diversity in their classrooms because they may fear that their own authority and power as teachers might be undermined (p. 30). Increasingly, feminist educators are recognizing that they are existing within societal and educational systems which work against the total amelioration of their power. As such, they are turning from the elimination or minimizing of power in classrooms to an examination of its many and varied forms.

Illuminating power in teaching and learning

Rather than eliminating power relations in teaching and learning in formal settings, then, feminist educators have turned toward illuminating those power relations. This attention has highlighted the myriad power dichotomies that are asserted within our educational systems, as well as the dynamic results of their interactions. Feminist educators are continuing their in-depth consideration and definition of the sources and enactments of power in educational settings.

First and foremost, teachers' own power in classrooms has been scrutinized. Are teachers powerful purely by the nature of their positions within academic settings? How does gender interact with power that is "granted" by institutional norms and behaviors? How do institutions mitigate the power of teachers? How do students mitigate the power of teachers? How do other identity characteristics of teachers and students create dynamic power differentials? And how do those dynamics affect teaching and learning?

These questions have been addressed by many feminist educators and are far from being settled. bell hooks (1996; orig. 1989) suggests that feminists who work in academic settings must always be aware of the ways in which the power of their practices are implicated by the contexts in which their actions are empowered. In her words, "As institutional structures impose values, modes of thought, ways of being on our consciousness, those of us who work in academic settings often unwittingly become engaged in the production of a feminist theory that aims to create a new sphere of theoretical elitism" (p. 57-58). Correspondingly, Kathleen Weiler (1988) agrees that institutional structures affect teachers' practices and power, but posits further that students' identities and interactions affect teachers as well. In her words, "The power of dominant structures is expressed not only in the institutional structure of the school, but is brought into the classroom itself in the consciousness and lived histories of students" (p. 124). Through an examination of their own power, many feminist teachers move to highlight, or illuminate, the power relations that allow for certain discourses to act and interact differently in higher education and in their specific classrooms.

Power is seen as being both repressive and productive in some feminist, critical and poststructural pedagogies (Gore, 1993). As such, feminist analyses of power in educational settings seek to respond to questions that consider what power allows and disallows in feminist classrooms and how that process occurs. For example, Marilyn Frye (1983) suggests that women, whether teachers or students, perform similar functions in that they serve men and men's interests (p. 9). Mimi Orner (1992) urges feminist educators to consider the ways in which they relate to students and youth as "Other" than themselves, as positioned in an oppositional stance, as not themselves (p. 76). Through these challenges, feminist scholars and educators have continued their efforts to illuminate power relations within their classrooms. This illumination in itself begins a disruption which, in turn, suggests new ways of understanding power and its effects in classrooms.

Disrupting and (re)presenting power through teaching and learning

"As soon as there is a power relation, there is a possibility of resistance. We can never be ensnared by power: we can always modify its grip in determinate conditions and according to a precise strategy" (Foucault, 1988; orig. 1977, p. 123).

Feminist educators have created a disruption of power relations. Their questioning of traditional power relations in classrooms and other social settings challenges what many students and teachers have come to expect. Further, their questioning becomes useful as an educational strategy in classroom settings, teaching about subject matter as well as educational and social processes and probing the assumptions that support related practices.

One way that many feminist educators suggest this disruption of commonly accepted practices should occur is through the empowerment of students (Bricker-Jenkins and Hooyman, 1986; hooks, 1994; Weiler, 1988). For example, Mary Bricker-Jenkins and Nancy Hooyman (1986) believe:

> Feminist concepts of power are within the context of a world view in which all are connected and therefore each responsible for the well-being of the whole; empowerment takes place as we observe feminist values in the process of seeking common ground and making common cause with each other. The view that power is limitless leads to a setting aside of competitiveness, which is based on win-lose, zero-sum thinking (p. 37).

And bell hooks (1984) insists as well that feminist education must insist on the power of women, rather on their powerlessness, in efforts to motivate them to action. In her words:

> Feminist movement would have had, and will have, a greater appeal for masses of women if it addresses the powers women exercise even as it calls attention to sexist discrimination, exploitation, and oppression. Feminist ideology should not encourage (as sexism has done) women to believe they are powerless. It should clarify for women the powers they exercise daily and show them ways these powers can be used to resist sexist domination and exploitation. Sexism has never rendered women powerless. It has either suppressed their strength or exploited it. Recognition of that strength, that power, is a step women together can take towards liberation (p. 93).

Through calls for empowerment of women and students, who have traditionally occupied oppressed roles, some feminist educators believe that they can effectively disrupt their classroom environments in educationally useful ways.

Still, questions about empowerment remain. How does empowerment happen in classrooms? Who does the empowering and who remains poised to be empowered? And, perhaps most importantly, to what end? While much literature on feminist teaching suggests a desire of teachers to empower their students through emancipatory practices, many scholars have expressed concerns about the establishment of unequal power statuses inherent in "empowering" relationships (Ellsworth, 1992; orig. 1989; Gore, 1993; Orner, 1992). By electing to use this method of disruption, feminist educators may come dangerously close to simply reinscribing old patterns of domination and creating new ones. For example, Mimi Orner (1992) insists:

> Calls for students to speak in the name of their own liberation and empowerment must be scrutinized. Educators concerned with changing unjust power relations must continually examine our assumptions about our own positions, those of our students, the meanings and uses of student voice, our power to call for students to speak, and our often unexamined power to legitimate and perpetuate unjust relations in the name of student empowerment (p. 77).

Jennifer Gore (1992) asserts further:

> In attempts to empower others we need to acknowledge that our agency has limits, that we might "get it wrong" in assuming we know what would be empowering for others, and that no matter what our aims or how we go about "empowering", our efforts will be partial and inconsistent (p. 63).

Empowerment as a strategy to disrupt traditional power relations are enacted is, as many critical and poststructural feminists have pointed out, both complex and potentially dangerous.

Questioning of commonly held norms and values by feminist teachers does not merely disrupt. Rather, it simultaneously reinscribes meaning, working to (re)create or (re)present meanings about societal and educational power relations. Therefore, traditional teacher/student, men/women, black/white, old/young dichotomies that largely proscribe the relations which are seemingly available to educators as they enter into educational interchanges with each other can change dramatically throughout this disruption.

Feminist teachers have long questioned the power relations that are inherent and unstable in their classroom experiences in the following ways:

- How can feminist teachers subvert norms and practices of patriarchy in their work?
- How does power shape the different experiences that class participants are able to construct?
- How does the location of feminist teaching in academic settings serve to reinscribe and support patriarchal structures?
- Is it possible to substantially reconstruct gender relations within feminist classrooms?

- What are the historical and political factors that have enabled certain educational discussions and decisions to occur?
- Who establishes the rules and norms for this environment? Are all persons who hold a stake in the outcomes of education represented and encouraged?

Feminist teachers strive to be aware of the ways in which they and other participants in educational settings are powerful. Further, they seek to understand the ways that power is subverted or altered within academic and other social settings. Most importantly, though, feminist teachers craft their roles as actors who have a responsibility to reconstruct educational environments so that they are conducive to teaching and learning for all involved.

CONTINUING THE CONVERSATION

"*A feminist practice can have its greatest force if, at the same time as it shifts the sands of an historical sedimentation, it leaves its own newly contoured landscape. That is, its own inevitable inaccuracy and lack of finality which must always show up wherever lines are drawn for the purpose of theoretical argument or for the more solid purpose of re-mapping an institution*" (Kamuf, 1990, p. 111).

"*Theories of knowledge, like theories about anything, are tools to organize, explain, and reconstruct our present and past experiences, and to be used in the service of our efforts to develop better knowledge and epistemic practices. Such theories, like all theories, emerge and evolve concomitantly with others and with the interests, values, projects, and practices that motivate and shape our work to generate them*" (Nelson, 1995, p. 45).

Our knowledge of the differences which constitute our (re)formed/(re)forming identities are embedded in power relations which continually renew themselves based on our own and others' desires and limitations. Feminist teaching practices, in their attention to knowledge, difference and power, present and represent the realities in which we can engage within educational and social settings. Through feminist teaching, feminists working in higher education have a great potential to influence the norms of teaching and learning that they, their colleagues, and students embrace or seek to disrupt.

Knowledge, difference and power are all implicated by and connected to each other. The web they create shifts regularly as social circumstances define the directions in which future spinning can occur. As such, they suggest many questions for future research. How do power relations in classrooms shape what is possible for teaching and learning? How do changes in the knowledge bases of our classrooms affect the degree to which we are able to work toward a disruption of inequitable power relations? The complex contexts in which we teach and learn enable us to spin enactments and understandings of knowledge, difference and power in intricate ways.

The questions throughout this work were generated within a certain context as well and are situated in a very specific time and place. My positioning as a new feminist teacher, among my many other identities, aspirations and ideologies, has urged me to consider questions that are relevant to my own developing experiences. For the most part, though, I have drawn on the questions of others as they have been presented in literature or in my own research on feminist teaching. Still, the web that I attempted to explore here would undoubtedly be seen and constructed differently if approached from another vantage point. It is my hope, though, that both feminist educators and others working in and seeking to understand higher education will find this examination of crucial issues in feminist teaching useful in crafting their practices and directions of inquiry.

References[8]

Acker, S. (1987). Feminist theory and the study of gender and education. *International Review of Education* 33: 419-435.

Alarcon, N. (1990). The theoretical subject(s) of *This bridge called my back* and Anglo-American feminism. In G. Anzaldua (ed.), *Making Face, Making Soul, Haciendo Caras: Creative and Critical Perspectives by Feminists of Color*. San Francisco: Aunt Lute Books.

Alcoff, L. (1988). Cultural feminism versus poststructuralism: The identity crisis in feminist theory. In E. Minnich, J. O'Barr, and R. Rosenfeld (eds.), *Reconstructing the Academy: Women's Education and Women's Studies*. Chicago: University of Chicago.

Anzaldua, G. (1990). *Making face, making soul, haciendo caras: Creative and critical perspectives by feminists of color*. San Francisco: Aunt Lute Books.

Apple, M. W. (1993). Between moral regulation and democracy: The cultural contradictions of the text. In C. Lankshear and P. L. McLaren (eds.), *Critical Literacy: Politics, Praxis, and the Postmodern*. Albany, NY: State University of New York.

Banks, A., Billings, D., and Tice, K. (1996; orig. 1993). Appalachian studies and postmodernism. In M. F. Rogers (ed.), *Multicultural Experiences, Multicultural Theories*. New York: McGraw-Hill.

Bar On, B.A. (1992). The feminist sexuality debates and the transformation of the political. *Hypatia* 7 (4): 45-58.

Barber, B. R. (1992). *An Aristocracy of Everyone: The Politics of Education and the Future of America*. New York: Oxford University.

Barbezat, D. (1988). Gender differences in the academic reward system. In D. W. Breneman and T. I. K. Youn (eds.), *Academic labor: Markets and careers*. New York: Taylor and Franklin.

Bee, B. (1993). Critical literacy and the politics of gender. In C. Lankshear and P. L. McLaren (eds.), *Critical Literacy: Politics, Praxis, and the Postmodern* . Albany, NY: State University of New York.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R., and Tarule, J. M. (1986). *Women's Ways of Knowing: The Development of Self, Voice, and Mind*. USA: Basic Books.

Bennett, R. S. (1991). Empowerment=Work over time: Can there be feminist pedagogy in the sport sciences? *Journal of Physical Education, Recreation, and Dance* 62: 62-67, 75.

[8] This reference list contains works that were reviewed for, but not cited in, this chapter.

Bezucha, R. J. (1985). Feminist pedagogy as a subversive activity. In M. Culley and C. Portuges (eds.), *Gendered Subjects: The Dynamics of Feminist Teaching*. Boston: Routledge and Kegan Paul.

Bloland, H. G. (1995). Postmodernism and higher education. *Journal of Higher Education* 66 (5): 521-559.

Boxer, M. J. (1985). Women's studies, feminist goals, and the science of women. In B. M. Solomon (ed.), *In the Company of Educated Women: A History of Women and Higher Education in America*. New Haven, CT: Yale University.

Brady, J. and Hernandez, A. (1993). Feminist literacies: Toward emancipatory possibilities of solidarity. In C. Lankshear and P. L. McLaren (eds.), *Critical Literacy: Politics, Praxis, and the Postmodern*. Albany, NY: State University of New York.

Bricker-Jenkins, M., and Hooyman, N. (1986). Feminist pedagogy in education for social change. *Feminist Teacher* 2 (2): 36-42.

Brown, J. (1992). Theory or practice--What exactly is feminist pedagogy? *Journal of General Education* 41: 51-63.

Bunch, C., and Pollack, C. (1983). *Learning Our Way: Essays in Feminist Education*. Trumansburg, NY: Crossing.

Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.

Calhoun, C. (1995). The gender closet: Lesbian disappearance under the sign "women." *Feminist Studies* 21 (1): 7-34.

Card, C. (1991). The feistiness of feminism. In *Feminist Ethics*. Lawrence, KS: University of Kansas.

Chamberlain, M. K. (1988). *Women in Academe: Progress and Prospects*. New York: Russell Sage Foundation.

Childers, M., and hooks, b. (1990). A conversation about race and class. In M. Hirsch and E. F. Keller (eds.), *Conflicts in Feminism*. New York: Routledge.

Collins, P. H. (1991). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. New York: Routledge, Chapman and Hall.

Collins, P. H. (1996; orig. 1986). Learning from the outsider within: The sociological significance of Black feminist thought. In M. F. Rogers (ed.), *Multicultural Experiences, Multicultural Theories*. New York: McGraw-Hill.

Culley, M., and Portuges, C. (1985). *Gendered Subjects: The Dynamics of Feminist Teaching*. Boston: Routledge and Kegan Paul.

D'Souza, D. (1992). *Illiberal Education: The Politics of Race and Sex on Campus*. New York: Vintage Books.

Damrosch, D. (1995). *We Scholars: Changing the Culture of the University*. Cambridge, MA: Harvard University.

De Lauretis, T. (1990). Upping the anti (sic) in feminist theory. In M. Hirsch and E. F. Keller (eds.), *Conflicts in Feminism*. New York: Routledge.

Dunn, K. (1993). Feminist teaching: Who are your students? *Women's Studies Quarterly* 21 (3and4): 39-45.

Dworkin, A. (1992; orig. 1983). The politics of intelligence. In E. Frazer, J. Hornsby, and S. Lovibond (eds.), *Ethics: A feminist reader*. Cambridge, MA: Blackwell.

Ellsworth, E. (1992; orig. 1989). Why doesn=t this feel empowering?: Working through the repressive myths of critical pedagogy. In C. Luke and J. Gore (eds.), *Feminisms and Critical Pedagogy*. New York: Routledge.

Firestone, S. (1993). The dialectic of sex. In S. Jackson (ed.), *Women's Studies: Essential Readings*. New York: New York University.

Firing Line. (1996). Resolved: Affirmative action should be terminated. Video production.

Fiske, E. (1992). Gender issues in the college classroom. In P. S. Rothenberg (ed.), *Race,*

Class, and Gender in the United States. New York: St. Martin's.
Flax, J. (1993). *Disputed Subjects: Essays on Psychoanalysis, Politics and Philosophy.* New York: Routledge.
Flax, J. (1996; orig. 1979). Women do theory. In M. F. Rogers (ed.), *Multicultural Experiences, Multicultural Theories*. New York: McGraw-Hill.
Foucault, M. (1988; orig. 1977). Power and sex. In L. D. Kritzman (ed.), *Michel Foucault: Politics, Philosophy and Cultures: Interviews and Other Writings 1977-1984*. New York: Routledge.
Foucault, M. (1978). *The History of Sexuality: An Introduction*. (Vol. 1). New York: Random House.
Fraser, N., and Nicholson, L. (1988). Social criticism without philosophy: An encounter between feminism and postmodernism. In A. Ross (ed.), *Universal Abandon: The Politics of Postmodernism*. Minneapolis: University of Minnesota.
Freccero, C. (1990). Notes of a post-sex wars theorizer. In M. Hirsch and E. F. Keller (eds.), *Conflicts in Feminism*. New York: Routledge.
Freire, P. (1990). *Pedagogy of the Oppressed*. New York: Continuum.
Frye, M. (1992). *Willful Virgin: Essays in Feminism*. Freedom, CA: Crossing Press.
Frye, M. (1983). *The Politics of Reality: Essays in Feminist Theory.* Freedom, CA: Crossing Press.
Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University.
Ginzberg, R. (1991). Philosophy is not a luxury. In *Feminist Ethics*. Lawrence, KS: University of Kansas.
Giroux, J. B. (1989). Feminist theory as pedagogical practice. *Contemporary Education* 61 (1): 6-10.
Giroux, H. (1993). Literacy and the politics of difference. In C. Lankshear and P. L. McLaren (eds.), *Critical Literacy: Politics, Praxis, and the Postmodern*. Albany, NY: State University of New York.
Gore, J. (1992). What we can do for you! What *can* 'we' do for 'you'?: Struggling over empowerment in critical and feminist pedagogy. In C. Luke and J. Gore (eds.), *Feminisms and Critical Pedagogy.* New York: Routledge.
Gore, J. (1993). *The Struggle for Pedagogies: Critical and Feminist Discourses as Regimes of Truth*. New York: Routledge.
Grumet, M. R. (1991). The politics of personal knowledge. In C. Witherell and N. Noddings (eds.), *Stories Lives Tell: Narrative and Dialogue in Education*. New York: Teachers College.
Gumport, P. J. (1990). Feminist scholarship as a vocation. *Higher Education* 20: 231-243.
Hartmann, H. (1993). The unhappy marriage of Marxism and feminism. In Stevi Jackson (ed.), *Women's Studies: Essential Readings*. New York: New York University.
Heald, S. (1989). The madwoman out of the attic: Feminist teaching in the margins. *Resources for Feminist Research* 18: 22-26.
Hensel, N. (1991). *Realizing Gender Equality in Higher Education: The Need to Integrate Work/Family Issues*. ASHE-ERIC Higher Education Report No. 2. Washington, D. C.: The George Washington University, School of Education and Human Development.
Hirsch, M., and Keller, E. F. (1990). Conclusion: Practicing conflict in feminist theory. In *Conflicts in Feminism*. New York: Routledge.
Holland, D. C., and Eisenhart, M. A. (1990). *Educated in Romance: Women, Achievement, and College Culture*. Chicago: University of Chicago.
hooks, bell. (1994). *Teaching to Transgress: Education as the Practice of Freedom*. New York: Routledge.
hooks, bell. (1996; orig. 1989). Feminist theory: A radical agenda. In M. F. Rogers (ed.),

Multicultural Experiences, Multicultural Theories. New York: McGraw-Hill.
hooks, bell. (1984). *Feminist Theory from Margin to Center.* Boston, MA: South End.
Howe, F. (1977). *Seven Years Later: Women's Studies Programs in 1976: A Report of the National Advisory Council on Women's Educational Programs.* New York: State University of New York.
Hull, G. T. and Smith, B. (1982). Introduction: The politics of black women's studies. In G. T. Hull, P. B. Scott, and B. Smith (eds.), *All the Women are White, All the Blacks are Men, But Some of Us are Brave: Black Women's Studies* (pp. xvii - xxxi). New York: Feminist Press.
Jipson, J., Munro, P., Victor, S., Froude-Jones, K., Freed-Rowland, G. (1995). *Repositioning Feminism and Education: Perspectives on Educating for Social Change.* Westport, CT: Bergin and Garvey.
Kamuf, P. (1990). Replacing feminist criticism. In M. Hirsch and E. F. Keller (eds.), *Conflicts in Feminism.* New York: Routledge.
Kelly, G. P., and Korsmeyer, C. (1991). Feminist scholarship and the American Academy. In G. P. Kelly and S. Slaughter (eds.), *Women's Higher Education in Comparative Perspective.* Netherlands: Kluwer Academic.
Kenway, J. and Modra, H. (1992; orig. 1989). Feminist pedagogy and emancipatory possibilities. In C. Luke and J. Gore (eds.), *Feminisms and Critical Pedagogy.* New York: Routledge.
Kerr, C. (1995). *The Uses of the University.* Cambridge, MA: Harvard University.
King, K. (1990). Producing sex, theory, and culture: Gay/straight remappings in contemporary feminism. In M. Hirsch and E. F. Keller (eds.), *Conflicts in Feminism.* New York: Routledge.
Kolodny, A. (1988). Dancing between left and right: Feminism and the academic minefield in the 1980s. *Feminist Studies* 14 (3): 453-466.
Lather, P. (1991). *Getting Smart: Feminist Research and Pedagogy With/In the Postmodern.* New York: Routledge.
Lazreg, M. (1990). Feminism and difference: The perils of writing as a woman on women in Algeria. In M. Hirsch and E. F. Keller (eds.), *Conflicts in Feminism.* New York: Routledge.
Lechte, J. (1994). *Fifty Key Contemporary Thinkers: From Structuralism to Postmodernity.* New York: Routledge.
Lee, P. A. (1994). To dance one's understanding. *Educational Leadership* 51: 81-83.
Lewis, M. (1990). Interrupting patriarchy: Politics, resistance and transformation in the feminist classroom. *Harvard Educational Review,* 60 (4): 467-488.
Lewis, M. G. (1993). *Without a Word: Teaching Beyond Women's Silence.* New York: Routledge.
Lorde, A. (1992). Age, race, class, and sex: Women redefining difference. In P. S. Rothenberg (ed.), *Race, Class, and Gender in the United States.* New York: St. Martin's.
Lorde, A. (1990). I am your sister: Black women organizing across sexualities. In G. Anzaldua (ed.), *Making Face, Making Soul, Haciendo Caras: Creative and Critical Perspectives by Feminists of Color.* San Francisco: Aunt Lute Books.
Luke, C. (1992). Feminist politics in radical pedagogy. In C. Luke and J. Gore (eds.) *Feminisms and Critical Pedagogy.* New York: Routledge.
Luke, C., and Gore, J. (1992). *Feminisms and critical pedagogy.* New York: Routledge.
Maher, F. A. (1985). Classroom pedagogy and the new scholarship on women. In M. Culley and C. Portuges (eds.), *Gendered Subjects: The Dynamics of Feminist Teaching.* Boston: Routledge and Kegan Paul.
Maher, F. A. (1987). Inquiry teaching and feminist pedagogy. *Social Education* 51 (3): 186-192.

Maher, F. A., and Tetreault, M. K. T. (1994). *The Feminist Classroom.* New York: Basic Books.

McLaren, P. L., and Lankshear, C. (1993). Critical literacy and the postmodern turn. In C. Lankshear and P. L. McLaren (eds.), *Critical Literacy: Politics, Praxis, and the Postmodern.* Albany, NY: State University of New York.

Middleton, S. (1993). *Educating Feminists: Life Histories and Pedagogy.* New York: Teachers College.

Middleton, S. (1995). Doing feminist educational theory: A post-modernist perspective. *Gender and Education* 7 (1): 87-100.

Minnich, E., O'Barr, J., and Rosenfeld, R. (1988). *Reconstructing the Academy: Women's Education and Women's Studies.* Chicago: University of Chicago.

Minow, M. (1990). Adjudicating differences: Conflicts among feminist lawyers. In M. Hirsch and E. F. Keller (eds.), *Conflicts in Feminism.* New York: Routledge.

Morton, D. (1996; orig.1993). The politics of queer theory in the (post) modern moment. In M. F. Rogers (ed.), *Multicultural Experiences, Multicultural Theories.* New York: McGraw-Hill.

Nelson, L. H. (1995). The very idea of feminist epistemology. *Hypatia* 10 (3): 31-49.

Nicholson, C. (1995). Postmodern feminisms. In M. Peters (ed.), *Education and the Postmodern Condition.* Westport, CT: Bergin and Garvey.

Nicholson, C. (1989). Postmodernism, feminism, and education: The need for solidarity. *Educational Theory* 39 (3): 197-205.

O'Barr, J. and Wyer, M. (1992). *Engaging Feminism: Students Speak Up and Speak Out.* Charlottesville, VA: University Press of Virginia.

Orner, M. (1992). Interrupting the calls for student voice in 'liberatory' education: A feminist poststructuralist perspective. In C. Luke and J. Gore (eds.), *Feminisms and Critical Pedagogy.* New York: Routledge.

Palmer, P. J. (1983). *To Know As We Are Known: A Spirituality of Education.* New York: HarperCollins.

Pearson, C. S., Shavlik, D. L, and Touchton, J. G. (1989). *Educating the Majority: Women Challenge Tradition in Higher Eeducation.* New York: American Council on Education, Macmillan.

Pelikan, J. (1992). *The Idea of the University: A Reexamination.* New Haven, CT: Yale University.

Pounder, D. G. (1989). The gender gap in salaries of educational administration professors. *Educational Administration Quarterly* 25 (2): 181-201.

Rakow, L. F. (1992). Gender and race in the classroom: Teaching way out of line. *Feminist Teacher* 6 (1): 10-13.

Regan, H. B. (1990). Not for women only: School administration as a feminist activity. *Teachers College Record* 91 (4): 565-577.

Robinson, L. H. (1973). *Women's Studies: Courses and Programs for Higher Education.* Washington, D. C.: American Association for Higher Education.

Rockhill, K. (1993). (Dis)connecting literacy and sexuality: Speaking the unspeakable in the classroom. In C. Lankshear and P. L. McLaren (eds.), *Critical Literacy: Politics, Praxis, and the Postmodern.* Albany, NY: State University of New York.

Ropers-Huilman, B. (1996a). Negotiating knowledge and knowing: Philosophies of teaching and learning in feminist classrooms. ERIC Clearinghouse on Higher Education (ED 387 031).

Ropers-Huilman, B. (1996b). Reframing resistance: Deconstructing feminist teaching discourse in higher education. Paper presented at the annual meeting for National Women's Studies Association, Saratoga Springs, New York, June, 1996.

Ropers-Huilman, B. (in press). *Feminist Teaching in Theory and Practice: Situating*

Power and Knowledge in Poststructural Classrooms. New York: Teachers College Press.

Rudolph, F. (1990; orig. 1962). *The American College and University: A History.* Athens, GA: University of Georgia.

Sadker, M., and Sadker, D. (1994). *Failing at Fairness: How America's Schools Cheat Girls.* New York: Macmillan.

Sawicki, J. (1991). *Disciplining Foucault: Feminism, Power, and the Body.* New York: Routledge.

Schniedewind, N. (1983). Feminist values: Guidelines for teaching methodology in Women's Studies. *Radical Teacher* 18: 25-28.

Schubert, J. D. (1995). From a politics of transgression toward an ethics of reflexivity: Foucault, Bourdieu, and academic practice. *American Behavioral Scientist* 38 (7): 1003-1017.

Schuster, M. R., Van Dyne, S. R. (1985). *Women's Place in the Academy: Transforming the Liberal Arts Curriculum.* Totowa, NJ: Rowman and Allanheld.

Scott, J. W. (1990). Deconstructing equality-versus-difference: Or, the uses of poststructuralist theory for feminism. In M. Hirsch and E. F. Keller (eds.), *Conflicts in Feminism.* New York: Routledge.

Shor, I. (1993). Education is politics: Paulo Freire=s critical pedagogy. In P. McLaren and P. Leonard (eds.), *Paulo Freire: A Critical Encounter.* New York: Routledge.

Shrewsbury, C. M. (1993). What is feminist pedagogy? *Women's Studies Quarterly* 21 (3-4): 8-15.

Simeone, A. (1987). *Academic Women: Working Towards Equality.* South Hadley, MA: Bergin and Garvey.

Smith, B. (1990). Racism and Women's Studies. In G. Anzaldua (ed.), *Making Face, Making Soul, Haciendo Caras: Creative and Critical Perspectives by Feminists of Color.* San Francisco: Aunt Lute Books.

Smith, L. (1994). Secret basketball: One problem with the student-centered classroom. *Feminist Teacher* 8 (1): 16-19.

Smith, P. (1990). *Killing the Spirit: Higher Education in America.* New York: Penguin.

Snitow, A. (1990). A gender diary. In M. Hirsch and E. F. Keller (eds.), *Conflicts in Feminism.* New York: Routledge.

Statham, A, Richardson, L., and Cook, J. A. (1991). *Gender and University Teaching.* Albany, NY: State University of New York.

Stimpson, C. R. (1975). The new feminism and women's studies. In Editors of Change (eds.), *Women on Campus: The Unfinished Liberation.* New Rochelle, New York: Change Magazine.

Stimpson, C. (1986). *Women's Studies in the United States.* New York: Ford Foundation.

Stitt, B. A. (1988). *Building Gender Fairness in Schools.* Carbondale, IL: Southern Illinois University.

Stoner, M. R. (1994). Rethinking feminist concerns: The evolution of feminist thought. *American Behavioral Scientist* 37 (8).

Swartzlander, S., Pace, S., Stamler, V. L.. (1993, February 17). The ethics of requiring students to write about their personal lives. *The Chronicle of Higher Education.*

Sypnowich, C. (1996; orig. 1993). Some disquiet about "difference." In M. F. Rogers (ed.), *Multicultural Experiences, Multicultural Theories.* New York: McGraw-Hill.

Thompson, A., and Gitlin, A. (1995). Creating spaces for reconstructing knowledge in feminist pedagogy. *Educational Theory* 45 (2): 125-150.

Thorne, B., Warren, C., and Geller. L. (1994). A feminist regrounding of sexuality and intimacy. *American Behavioral Scientist* 37 (8): 1042-1057.

Tong, R. (1989). *Feminist Thought: A Comprehensive Introduction.* Boulder, CO: Westview.

Uttal, L. (1990). Inclusion without influence: The continuing tokenism of women of color. In G. Anzaldua (ed.), *Making Face, Making Soul, Haciendo Caras: Creative and Critical Perspectives by Feminists of Color.* San Francisco: Aunt Lute Books.

Vaid, U., Wolf, N., Steinem, G., and hooks, b. (1993, September/October). Let's get real about feminism: The backlash, the myths, the movement. *Ms.* 4 (2): 34-43.

Walkerdine, V. (1993). The mastery of reason. In S. Jackson (ed.), *Women's Studies: Essential Readings.* New York: New York University.

Washington, M. (1996). The minority student in college: A historical analysis. In C. S. V. Turner, M. Garcia, A. Nora, and L. I. Rendon (eds.), *Racial and Ethnic Diversity in Higher Education.* Needham Heights, MA: Simon and Schuster.

Weedon, C. (1987). *Feminist Practice and Poststructuralist Theory.* Oxford: Basil Blackwell.

Weiler, K. (1988). *Women Teaching for Change: Gender, Class, and Power.* New York: Bergin and Garvey.

Yamato, G. (1990). Something about the subject makes it hard to name. In G. Anzaldua (ed.), *Making Face, Making Soul, Haciendo Caras: Creative and Critical Perspectives by Feminists of Color.* San Francisco: Aunt Lute Books.

Young-Bruehl, E. (1988). The education of women as philosophers. In E. K. Minnich, J. O'Barr, and R. Rosenfeld (eds.), *Reconstructing the Academy: Women's Education and Women's Studies.* Chicago: University of Chicago.

Women and Minorities in Higher Education[1]

Larry L. Leslie and Ronald L. Oaxaca
University of Arizona

The underutilization of minorities and women in science and engineering is one of the most vexing problems of U.S. higher education. Although women have nearly reached or even exceeded parity with men as entrants into numerous occupational fields, their numbers in many science and engineering academic programs and occupations remain substantially below their distributions in the population (National Research Council, 1991, 1994; Seymour and Hewitt, 1994; Commission on Professionals, 1994). Although some progress has been made, the participation of minorities in science and engineering is considerably behind that of women (Commission on Professionals, 1994; National Research Council, 1994).

We say "vexing" because progress appears to be modest despite enormous efforts and numerous successes in understanding and addressing the underlying causes of the problems. In our extensive review of the related literature, we identified no less than 120 empirical and theoretical undertakings related directly to the issue. Further, during the 30 months of the project, more than a score of important new papers came to our attention (e.g., Ethington, 1995; Paulson et al., 1995; Ross, Volkvein, and Vogt, 1995; Yaeger, 1995). In addition to the usual commentary-type articles, this substantial literature ranges from studies that consider specific questions by examining small samples in single institutions to major national investigations of all the conceivable issues bearing on the larger question.

In this chapter we present not only the results of our own empirical attempts to understand the reasons for female and minority underrepresentation in science and engineering, but also the results of our ambitious attempt to synthesize the work completed by others. Although a few writers have undertaken such syntheses, all have focused on either a single theoretical perspective (e.g., Holland and

[1]This material is based upon work sponsored by the National Science Foundation under Grant Number SBR-9311351. The Government has certain rights in this material. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.}

J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*

Eisenhart, 1990; Lent and Hackett, 1987) or on a particular component of the educational process, such as the college years (e.g., Castle, 1993; Mow and Nettles,1990).

Our effort has been as broad as possible. Although our empirical work has concentrated on the higher education years, our data set (the Cooperative Institutional Research Program [CIRP]) includes substantial information from the pre college years, taken at the time of college entrance, and extends through undergraduate and graduate education and employment. Our treatment herein follows a "life-sequence" approach, and incorporates three major theories or concepts that we identified in the literature.

Our organization is somewhat unusual: We incorporate our own findings with those from the review; that is, we use the review as our base and add what we have found to what previously has been known. After briefly describing our methods, as much as feasible we arrange our treatment starting with the early years of life, progressing on through school and college years, and finishing nine years after college matriculation at a time when some individuals are employed and others are in graduate school. We conclude the chapter with some policy implications.

DATA AND MODELS

CIRP data are collected annually from first-year students at the time of their entry into college. Periodically, follow-up surveys are conducted approximately nine years after matriculation, thereby permitting analysis of the interaction of collegiate and post-collegiate experiences with student characteristics and conditions. (Full details of the samples are described in the annual CIRP reports published by the Higher Education Research Institute at UCLA, e.g., Higher Education Research Institute, 1995.)

The CIRP data are almost entirely categorical as opposed to continuous. This necessitated our taking an exclusively limited dependent variable approach to the empirical analysis. We estimated binomial logit, multinomial logit, and ordered logit models by maximum likelihood methods. The definitions of the variables used are reported in the appendix. Because the estimated parameters of these models are not always easily interpreted, we focus our discussion of the empirical results on the estimated marginal effects of the explanatory variables (evaluated at the sample means) on the outcome probabilities. Because the outcome probabilities must sum to one, the marginal effects must sum to zero. Since the explanatory variables are typically dummy (indicator) variables, the notion of a marginal effect requires special consideration. We calculate the marginal effects by looking at the incremental effects of changing the values of the dummy variable from their sample means. One interpretation is that the sample means represent the probabilities of selecting someone with the given characteristics. We then estimate the effects on the outcome probabilities of marginal changes in the probabilities of drawing individuals with the given characteristics. Although a particular marginal effect may be relatively small, this does not imply that the estimated coefficients on the given variable are statistically insignificant in the underlying logit model. Further, it must be remembered that these are marginal effects; that is, the contributions of other variables are held constant. For example, it will be noted that having a parent who has a science or engineering occupation adds to the likelihood that one will major in science or engineering, quite aside from other aspects of that occupation, such as a income level, and quite aside from one's personal traits, such as high school rank and high school preparation in science and engineering.

Where number of cases in the sample permit, we estimate separate models for white males, white females, black males, black females, Hispanic males, and Hispanic females. Most often our analysis is limited to white males versus white females.

Our presentation of tables and data in this chapter is limited. Full information is contained in our report to NSF. The following text illustrates the nature of results, using only seven of more than forty tables in our NSF report.

For purposes of analyzing factors that determine the probabilities of selecting a first choice from among competing college majors in the "freshman" year, we aggregated choices into five broad categories of college major: physical sciences/ engineering, biological sciences, liberal arts, business, and all others. The underlying choice framework is represented by a multinomial logit model. The final estimated model included seven sets of explanatory variables, with each set consisting of one or more dummy variables. These variable sets are high school rank, self-rating of math and natural science preparation for college, the presence of a college prep program in high school, parental occupation in the science and engineering area, parental income, expectations regarding sources of college financing, and future marital plans. The sample size was 9,628. We estimated a pooled model and controlled for gender and ethnicity with dummy variables. We also estimated separate models for each of the six gender/ethnic categories.

First we discuss findings from the pooled sample. The estimated multinomial logit model correctly predicted 46.7% of the first choices of college major. The estimated marginal effects for the pooled sample are reported in Table 1. The magnitude of the effects may be evaluated by recalling that theoretically the marginal effects must lie between -1.0 and +1.0; but also note that the variables may or may not be statistically significant. As reflected by the negative signs (-), holding constant other factors, white females (WFEMALE) and Hispanic females (HFEMALE) are less likely (than white males) to select science and engineering or business as their first choice major and are more likely to select a major in liberal arts. For example, if the probability that a student is a Hispanic female were to increase by 10 percentage points at the mean, the probability of selecting a major in physical science or engineering would fall by 1.5 percentage points. Black females (BFEMALE) also are less likely (than white males) to select their first choice from physical science/engineering or business; however, black females are more likely to select a major in the biological sciences as well as the liberal arts. Black males (BMALE) were less likely (than white males) to select their first choice from science and engineering and more likely to select business. Finally, Hispanic males (HMALE) were less likely to select biological science or business but were more likely to select physical sciences/engineering.

TABLE 1. Marginal Effects for the First Choice of College Major: Total Sample (CIRP)

Variable	*PhSci/Eng*	*BioSci*	*Lib Arts*	*Business*	*Other*
WFEMALE	-0.1115	-0.0155	0.1628	-0.0552	0.0195
BFEMALE	-0.1144	0.0216	0.1089	-0.0298	0.0137
HFEMALE	-0.1475	-0.0298	0.1686	-0.0032	0.0119
BMALE	-0.0217	-0.0367	0.0130	0.0251	0.0203
HMALE	0.0028	-0.0156	0.0231	-0.0493	0.0390
HSRANK4	0.0526	0.0589	0.0074	-0.0391	-0.0799
HSRANK3	-0.0050	0.0236	0.0059	-0.0024	-0.0221
HSPRIVUS	-0.0222	0.0195	0.0392	0.0059	-0.0424
PREPSE	0.0789	0.0782	-0.1615	0.0138	-0.0094
CPREPSCH	-0.0064	0.0015	0.0693	-0.0433	-0.0210
PAROCSE	0.0506	0.0283	-0.0257	-0.0373	-0.0159
PARINC1	-0.0075	-0.0390	-0.0456	0.0312	0.0610
PARINC2	0.0086	-0.0144	-0.0504	0.0035	0.0526
PARINC3	0.0110	-0.0367	-0.0476	0.0094	0.0638
PARINC4	0.0270	-0.0188	-0.0354	-0.0022	0.0295
PARINC5	0.0079	-0.0088	-0.0226	0.0025	0.0210
PARINC6	0.0101	-0.0216	-0.0072	-0.0025	0.0212
PARINC8	-0.0112	0.0007	0.0472	-0.0277	-0.0090
PARINC9	-0.0102	0.0205	0.0162	-0.0227	-0.0038
PARINC10	-0.0114	-0.0050	0.0623	-0.0048	-0.0411
PARINC11	-0.0701	0.0483	0.0310	-0.0132	0.0041
PARINC12	-0.0493	0.0362	0.1194	-0.0731	-0.0333
WORKSAV	-0.0043	0.0010	-0.0193	0.0091	0.0135
PARENT	-0.0129	0.0015	0.0220	-0.0006	-0.0099
GRANTS	0.0163	-0.0008	0.0544	-0.0255	-0.0445
LOANS	-0.0117	0.0048	0.0415	-0.0218	-0.0127
FUTREMAR	0.0032	-0.0320	0.0072	0.0106	0.0112

Turning to the separately estimated multinomial logit models, we find the following predictive accuracy rates for first choice of college major: white males—37.5%, white females—55.0%, black males—40.4%, black females—51.7%, Hispanic males—51.9% and Hispanic females—56.7%. We report the separately estimated marginal effects in Tables 2 through 7. Our focus will be on the major differences, among groups, in the marginal effects of the determinants of first choice of college major as they pertain to selection of majors in science and engineering (physical sciences/engineering + biological sciences).

TABLE 2. Marginal Effects for the First Choice of College Major: White Males (CIRP)

Variable	*PhSci/Eng*	*BioSci*	*Lib Arts*	*Business*	*Other*
HSRANK4	0.0669	0.0972	0.0166	-0.0642	-0.1166
HSRANK3	-0.0158	0.0741	-0.0169	-0.0014	-0.0401
HSPRIVUS	-0.0452	0.0089	0.0553	0.0414	-0.0603
PREPSE	0.1022	0.0747	-0.1579	0.0102	-0.0291
CPREPSCH	-0.0122	0.0159	0.0621	-0.0360	-0.0299
PAROCSE	0.1073	0.0097	-0.0207	-0.0213	-0.0750
PARINC1	-0.0197	-0.0283	-0.0068	-0.0141	0.0689
PARINC2	0.0385	-0.0084	-0.0643	-0.0322	0.0663
PARINC3	0.0195	-0.0816	0.0231	-0.0219	0.0608
PARINC4	0.0484	-0.0364	-0.0588	-0.0140	0.0606
PARINC5	0.0037	-0.0127	-0.0106	-0.0087	0.0283
PARINC6	0.0399	-0.0289	-0.0447	0.0070	0.0266
PARINC8	-0.0168	0.0114	0.0375	-0.0503	0.0182
PARINC9	-0.0531	0.0081	0.0333	-0.0311	0.0428
PARINC10	-0.0533	0.0157	0.0586	-0.0003	-0.0206
PARINC11	-0.2316	0.0845	0.0773	0.0153	0.0546
PARINC12	-0.0923	0.0800	0.0970	-0.0942	0.0095
WORKSAV	0.0067	-0.0215	-0.0180	0.0057	0.0272
PARENT	-0.0372	0.0191	0.0341	-0.0164	0.0004
GRANTS	0.0190	0.0188	0.0426	-0.0571	-0.0234
LOANS	-0.0357	-0.0104	0.0664	-0.0191	-0.0012
FUTREMAR	-0.0105	-0.0063	-0.0173	0.0396	-0.0055

TABLE 3. Marginal Effects for the First Choice of College Major: White Females (CIRP)

Variable	*PhSci/Eng*	*BioSci*	*Lib Arts*	*Business*	*Other*
HSRANK4	0.0515	0.0211	0.0069	-0.0262	-0.0533
HSRANK3	0.0110	-0.0064	0.0290	-0.0011	-0.0326
HSPRIVUS	0.0020	0.0332	0.0096	-0.0149	-0.0299
PREPSE	0.0620	0.0856	-0.1724	0.0153	0.0094
CPREPSCH	-0.0041	-0.0219	0.0557	-0.0372	0.0074
PAROCSE	0.0235	0.0351	-0.0548	-0.0268	0.0231
PARINC1	-0.0016	0.0053	-0.0968	0.0371	0.0560
PARINC2	0.0140	0.0185	-0.0855	0.0205	0.0324
PARINC3	0.0170	-0.0273	-0.0799	0.0288	0.0615
PARINC4	0.0044	0.0018	-0.0192	0.0060	0.0070
PARINC5	-0.0013	0.0047	-0.0249	0.0118	0.0097
PARINC6	-0.0128	-0.0115	0.0318	-0.0178	0.0103
PARINC8	-0.0011	-0.0084	0.0530	-0.0183	-0.0253
PARINC9	0.0158	0.0330	0.0189	-0.0169	-0.0508
PARINC10	-0.0014	-0.0297	0.0919	0.0006	-0.0615
PARINC11	-0.0000	0.0470	0.0134	-0.0422	-0.0182
PARINC12	-0.0145	-0.0130	0.1343	-0.0384	-0.0684
WORKSAV	-0.0077	0.0179	-0.0222	0.0095	0.0025
PARENT	0.0018	-0.0109	0.0223	-0.0027	-0.0105
GRANTS	0.0198	-0.0096	0.0727	-0.0246	-0.0583
LOANS	-0.0044	0.0089	0.0159	-0.0222	0.0018
FUTREMAR	0.0138	-0.0508	0.0364	-0.0117	0.0124

TABLE 4. Marginal Effects for the First Choice of College Major: Black Males (CIRP)

Variable	*PhSci/Eng*	*BioSci*	*Lib Arts*	*Business*	*Other*
HSRANK4	0.0779	0.0587	-0.0606	0.0003	-0.0764
HSRANK3	0.0205	-0.0321	-0.0642	0.0128	0.0630
HSPRIVUS	-0.1006	0.0061	0.1537	-0.0267	-0.0325
PREPSE	0.0714	0.0550	-0.1167	0.0119	-0.0217
CPREPSCH	-0.0198	0.0324	0.0414	-0.0399	-0.0142
PAROCSE	0.1559	0.1367	0.7729	-1.3174	0.2518
PARINC1	-0.0207	-0.0831	0.0182	0.0346	0.0510
PARINC2	-0.0163	-0.0720	-0.0423	0.0686	0.0620
PARINC3	0.0267	-0.0156	-0.1644	0.0895	0.0637
PARINC4	0.1110	-0.0239	-0.1064	0.0604	-0.0411
PARINC5	0.1506	-0.1468	-0.1347	0.0787	0.0522
PARINC6	0.0699	-0.0278	-0.0871	0.0832	-0.0382
PARINC8	-0.0822	-0.0017	0.0732	0.0452	-0.0345
PARINC9	-1.1888	0.0786	0.5727	0.1570	0.3805
PARINC10	0.6319	0.4789	1.4123	-0.9142	-1.6089
PARINC11	0.0454	-0.1212	-1.2662	1.7006	-0.3586
PARINC12	-1.0878	0.2810	1.3762	-1.1246	0.5552
WORKSAV	-0.0298	0.0005	0.0028	-0.0081	0.0346
PARENT	-0.0141	0.0406	-0.0125	0.0099	-0.0240
GRANTS	0.0109	-0.0239	0.0053	-0.0020	0.0098
LOANS	-0.0136	0.0058	0.0821	0.0079	-0.0822
FUTREMAR	0.0093	-0.0219	-0.0075	0.0566	-0.0365

TABLE 5. Marginal Effects for the First Choice of College Major: Black Females (CIRP)

Variable	*PhSci/Eng*	*BioSci*	*Lib Arts*	*Business*	*Other*
HSRANK4	0.0340	0.0141	-0.0010	-0.0199	-0.0273
HSRANK3	0.0059	-0.0478	0.0188	-0.0016	0.0247
HSPRIVUS	-0.0181	-0.0113	0.0840	-0.0463	-0.0083
PREPSE	0.0384	0.0624	-0.0963	-0.0212	0.0167
CPREPSCH	0.0034	0.0146	0.1137	-0.0499	-0.0818
PAROCSE	-0.0008	0.1542	0.3809	-0.5504	0.0161
PARINC1	0.0117	-0.0850	-0.0186	0.0240	0.0679
PARINC2	0.0016	-0.0514	-0.0128	0.0015	0.0610
PARINC3	0.0021	-0.0249	0.0131	-0.0069	0.0166
PARINC4	-0.0062	-0.0165	-0.0380	0.0010	0.0597
PARINC5	0.0081	-0.0313	0.0448	-0.0155	-0.0061
PARINC6	-0.0023	-0.0290	-0.0469	-0.0165	0.0947
PARINC8	-0.0169	0.0319	0.0121	0.0332	-0.0602
PARINC9	-0.2265	0.1826	-0.0736	0.0163	0.1013
PARINC10	-0.2307	-0.0796	0.2068	0.0109	0.0926
PARINC11	-0.1007	-1.2704	3.4304	-0.3371	-1.7222
PARINC12	-0.1771	0.6029	2.0266	-0.4184	-2.0340
WORKSAV	-0.0013	0.0333	-0.0153	0.0071	-0.0238
PARENT	0.0157	-0.0348	-0.0023	0.0149	0.0065
GRANTS	0.0102	-0.0196	0.0523	0.0023	-0.0452
LOANS	0.0024	0.0210	0.0450	-0.0152	-0.0532
FUTREMAR	0.0005	-0.0705	0.0341	-0.0072	0.0431

TABLE 6. Marginal Effects for the First Choice of College Major: Hispanic Males (CIRP)

Variable	*PhSci/Eng*	*BioSci*	*Lib Arts*	*Business*	*Other*
HSRANK4	0.0298	0.0799	0.1363	-0.0278	-0.2182
HSRANK3	-0.0827	0.0518	0.0716	-0.0437	0.0029
HSPRIVUS	-0.0920	0.0057	0.0808	0.0154	-0.0099
PREPSE	0.1692	0.0494	-0.2723	0.0615	-0.0078
CPREPSCH	-0.0240	-0.0049	0.1212	-0.0091	-0.0832
PAROCSE	0.7993	-0.7711	2.2908	-0.4706	-1.8484
PARINC1	-0.0373	0.1529	-0.4767	0.1257	0.2355
PARINC2	0.0887	0.0878	-0.4160	0.0851	0.1543
PARINC3	0.0678	0.0813	-0.4833	0.0681	0.2662
PARINC4	0.1612	0.0486	-0.2587	0.0267	0.0222
PARINC5	0.1286	0.0862	-0.3481	-0.0059	0.1392
PARINC6	0.0898	0.1422	-0.1781	0.0488	-0.1027
PARINC8	-1.1509	0.2133	0.3853	0.1286	0.4237
PARINC9	-0.2571	-0.2186	-2.3000	-0.2000	2.9757
PARINC10	1.2539	-0.4247	-2.9292	-0.2203	2.3202
PARINC11	—	—	—	—	—
PARINC12	-0.2580	-0.2389	-2.4046	-0.1068	3.0083
WORKSAV	-0.1954	0.0344	0.0574	0.0440	0.0596
PARENT	-0.0906	0.0890	-0.0204	0.1005	-0.0786
GRANTS	0.0321	0.0138	0.1075	-0.0304	-0.1230
LOANS	0.0729	-0.0066	-0.0134	-0.0376	-0.0154
FUTREMAR	0.0324	0.0153	-0.1913	-0.0139	0.1575

TABLE 7. Marginal Effects for the First Choice of College Major: Hispanic Females (CIRP)

Variable	*PhSci/Eng*	*BioSci*	*Lib Arts*	*Business*	*Other*
HSRANK4	0.0008	0.0678	-0.0579	0.0534	-0.0641
HSRANK3	-0.0064	0.0234	0.0551	-0.0338	-0.0383
HSPRIVUS	0.0002	0.0190	-0.0250	0.0026	0.0032
PREPSE	0.0004	0.0486	-0.0078	0.0201	-0.0614
CPREPSCH	0.0006	0.0870	0.0843	-0.0693	-0.1026
PAROCSE	-0.0066	0.0952	0.4827	-1.1502	0.5789
PARINC1	-0.0013	-0.1383	0.3973	-0.1575	-0.1002
PARINC2	-0.0017	-0.1031	0.3803	-0.2359	-0.0396
PARINC3	-0.0015	-0.1576	0.2834	-0.2449	0.1206
PARINC4	-0.0010	-0.1288	0.5309	-0.2239	-0.1772
PARINC5	-0.0009	-0.1252	0.3477	-0.1976	-0.0240
PARINC6	-0.0071	-0.1439	0.2595	-0.1270	0.0184
PARINC8	0.0007	-0.8464	3.4810	-0.8563	-1.7791
PARINC9	—	—	—	—	—
PARINC10	—	—	—	—	—
PARINC11	—	—	—	—	—
PARINC12	-0.0055	-0.7933	3.8022	-1.0998	-1.9036
WORKSAV	-0.0001	0.0358	-0.0224	0.0168	-0.0302
PARENT	-0.0004	0.0121	0.1459	-0.0678	-0.0898
GRANTS	-0.0009	-0.0393	0.1436	-0.0089	-0.0946
LOANS	0.0000	0.0090	0.0853	-0.0961	0.0018
FUTREMAR	-0.0007	-0.0204	-0.0134	-0.0118	0.0463

Among the six gender and racial/ethnic groups, being in the top 25% of one's high school class had the largest positive effects on the selection of science and engineering for white males (Table 2). The effect of better than average preparation in math and natural science was the highest among Hispanic males (Table 6). A 10 percentage point increase in the probability of reporting better than average preparation increased the probability of selecting a major in science and engineering by 2.2 percentage points for Hispanic males and 1.7 percentage points for white males. Among males having a parent in a science and engineering occupation had the largest effect on the probability of selecting physical science and engineering for Hispanics. For example, increasing the probability of having a parent in a science and engineering occupation by 10 percentage points increases the probability of selecting a major in physical science and engineering by 8.0 percentage points for Hispanic males followed by 1.1 percentage points for white males. In the case of the biological sciences, blacks exhibited the largest positive effects of having a parent in science and engineering. A 10 percentage point increase in the probability of having a parent in science and engineering increased the probability of choosing a major in biological sciences by 1.5 percentage points for black females (Table 5) and 1.4 percentage points for black males (Table 4). For Hispanic males the effect was to lower the probability of selecting biological sciences by 7.7 percentage points; for Hispanic females the effect was a +1.0 percentage points (Table 7). Anticipation of reliance upon work, savings, or GI benefits (WORKSAV) had the largest effect on selection of a major in physical science and engineering for Hispanic males. This effect is negative and indicates that a 10 percentage point increase in the probability of relying upon work, savings, or GI benefits reduces the probability of selecting a major in physical science and engineering by 2.0 percentage points.

This illustrates the nature of the analysis, the form of the findings, and the basis for reporting results. The section that follows deals exclusively with results.

SYNTHESIS AND RESULTS

In preface we should note that space limitations required that we be selective in presenting results. Our decision was to concentrate on what we took to be the most powerful conceptual explanations for underrepresentation of women and minorities in science and engineering, necessitating the near-exclusion of much useful information, some of which is related, if not central, to the concepts themselves.[2] In this essay, brief mention is made of these related findings where the most important insights are to be gained.

[2]The list of (largely) omitted sections include variables related to the following: family, school and college personnel, instructional pedagogy, economic variables, and college environment and characteristics. The full report to NSF is available from the authors upon request.

The three concepts chosen were self-concept/self-efficacy, peer influence, and goal commitment, all of which are closely related. We begin the discussion with self-concept and self-efficacy, two closely connected abstractions that constitute an important body of research in psychology and social psychology. An individual's development of the first of these, self-concept, is fundamental to development of the latter, self-efficacy. The early development of self-concept bears importantly on the achievement of science outcomes for women and minorities later in life, i.e., through science self-efficacy. In turn one's self-concept/self-efficacy is impacted importantly by peers, all of which affect one's commitment to science and engineering.

1. Self-Concept and Self-Efficacy

Self-concept is perception of self; self-efficacy is belief in one's ability to perform a given behavior (Lent, Lopez, and Bieschke, 1991). The evidence in the literature is quite compelling that self-concept and self-efficacy affect strongly the achievement of science and engineering outcomes for women and although this literature is much more limited for minorities, there can be little doubt that the same is true for them.

Self-concept is hierarchical in nature. *General* self-concept has been measured and quantified, as have *Academic* self-concept and *Verbal* and *Mathematical* self-concepts, all of which have been shown to be interrelated. General self-concept is at the top of the self-concept hierarchy; positive, subordinate self-concepts generally contribute to a positive general self-concept although the former may be complementary rather than complimentary.

General self-concept is observable in very young children; subordinate self-concepts are not. Subordinate self-concepts and self-concept hierarchies begin to develop by about ages 5-8 (Marsh and Shavelson, 1985; Marion and Coladarci, 1993). General self-concept is quite stable but with age becomes situation-specific as the hierarchy is descended. Self-concept has been linked empirically to achievement (e.g., Shavelson et al., 1976). Math and verbal self-concepts are distinct and separate by late adolescence. These statements pertain to the overall population. What of gender differences?

Generally, in elementary school boys and girls do not vary significantly in math/science ability, confidence, or interest; however, many math/science gender differences are evident by the end of high school, with the junior high school years probably being transition years for most youth, but particularly girls (AAUW, 1991; Hyde, Fennema, and Lamon, 1990; Linn and Hyde, 1989; Meece et al., 1982[3]). It is instructive to note that these patterns are consistent with self-concept theory and with empirical evidence: the absence of self-concept differentiation in the early years and the distinctiveness and separateness

[3]This publication contains by far the most extensive extant literature review regarding gender-linked factors associated with academic choices although the focus is almost exclusively on mathematics. We draw heavily upon the Meece et al., review.

of math and verbal self-concept in later years.

Generally, self-efficacy, rather than self-concept, is employed to explain behavior related to science and engineering outcomes. This choice is supported by the definitional differences. Self-efficacy has a more specific contextual meaning than self-concept. As typically used, self-concept might be seen as the general conception of self one brings to personal decision making. It is the basis for forming one's sense of self-efficacy, which relates to specific decisions, such as career choices in science and engineering.

The theoretical and empirical literature regarding self-concept and especially self-efficacy is extremely useful in understanding specific science and engineering-related empirical findings about gender, and to some degree, racial/ethnic differences. This literature usually draws careful attention to the fundamental theoretical ideas.

Self-efficacy theory, which usually is credited to Bandura (1977), begins with the notion that human behavior is devised through cognitive processes, with performance-based experiences being most important. Learning about ourselves involves evaluating "differential consequences"; it involves acquiring and evaluating personal information. The elements of the learning process are four in number: We learn through "performance accomplishment," which is very important and is based upon mastery, success increasing mastery expectations and failure lowering it; "vicarious experience," which involves making social comparisons and is less powerful; "verbal persuasion," which is widely utilized but is of limited utility; and "emotional arousal." As knowledgeable readers about the problems of women and minorities in science and engineering will be aware, it is noteworthy to our discussion that the fourth element can be counterproductive, that emotional reactions can escalate and can lead to self-defeating behavior. This realization may be critical to our understanding of the issues addressed herein: Self-efficacy, when formulated, will be instrumental in determining whether one decides to cope with adversity, to what degree, and how persistently, as specific obstacles arise. Bandura observes significantly that the "strength of conviction" in one's self-efficacy will affect one's willingness to see a task or goal to completion and even whether one will make an effort:

> Weak expectations are easily extinguishable by disconfirming experiences, whereas individuals who possess strong expectations of mastery will persevere in their coping efforts despite disconfirming experiences.

Strong self-efficacy can turn threatening obstacles into events perceived as safe. Empirically, self-efficacy has been shown to predict performance accurately in 85% of tasks confronted (Bandura, 1977).

With this theory as a bridge, we may examine more closely human behavior as it pertains to science and engineering. On essentially all characteristics hypothesized to affect science and engineering-related educational and career choices, girls do not differ from boys in the early years, which are *prior to* the development of the distinctively different subordinate self-concepts; however, numerous, important self-concept/self-efficacy differences related to science and math are evident thereafter.

As a point of departure in examining gender differences related to science and engineering, ability differences between boys and girls in math and science are almost universally agreed to be minimal over all age ranges (e.g., Friedman, 1989; Orenstein, 1994).[4] Although some differences usually are noted in empirical work, analytical models almost uniformly find non-significance for ability measures when other factors, such as number of science and mathematics courses taken, are controlled. Even so, certain exceptions are instructive. Mathematical abilities do decline earlier and more steeply for girls, and at the high end of the ability continuum, differences clearly are noted; for example, males outnumber females 4:1 in the number of individuals earning more than 600 on the SAT-Math test (Benbow and Stanley, 1983). Yet, in spite of such differences, girls tend to earn somewhat higher grades in science and mathematics. Overall, the conclusion is that ability differences fail to account for important gender disparities in science and engineering and that we must look to other explanations.

Such differences are noted on a number of other measures. The number of *mathematics courses taken* almost invariably is the primary predictor of majoring and persistence in science and engineering (e.g., Astin and Astin, 1992). By the end of high school, boys substantially exceed girls by this measure. One's *expectations* of the probability of success in a given endeavor is known to be an important factor in formulating behavioral choices. Levels of expectations in science and math are lower for females than for males (e.g., Lent, Lopez, and Bieschke, 1991). There are several reasons for this.

For example, males rate the usefulness of science and math higher than do females, who even rate science and math as more useful for males than for themselves. Female ratings of science and mathematics usefulness decline for girls as they mature, a phenomenon that does not characterize the maturation of boys (Brush, 1980; Parsons and Adler, 1983; Sherman, 1980). No doubt female "other directedness" plays some part in this, that is the desire to please others. Linn and Hyde (1989) conclude that of all variables examined, only usefulness of math and science appear to affect persistence, and that this difference by gender has existed for many years.

Usefulness of science and math has been connected empirically to "valuing" science and math. Males value these subjects more than females (Betz and Hackett, 1983). The importance of this valuing to science-and-mathematics-linked academic and career choices is illustrated by analogy to the gender differences in the valuing of athletics: Since traditionally girls have valued athletics less than

---

[4]Eccles (1987) has published a review of this topic.

boys, it is hardly surprising that historically more boys than girls have engaged in athletics although, of course, this is changing.

It is similarly no surprise that boys voice a greater *interest* in and more positive *attitudes* toward science and mathematics than do girls although, again, the differences emerge only after the elementary school years (Betz and Hackett, 1983). Prior to adolescence, interest levels are essentially the same. In one study interest even surpassed ability in predicting self-efficacy in engineering, and self-efficacy was found to intervene to mediate effects of such factors as stress, coping, and gender and ethnicity on achievement (Hackett et al., 1992).

The same is true for *confidence* in science and math: Relatively and on average, women may lack such confidence (Manis et al., 1989). The disparity in favor of boys first appear in high school, and this relative confidence may not be supported by ability differences (Betz and Hackett, 1983). In turn confidence in science and mathematics has been connected to test-taking skills in math/science; for example, boys are more willing or able to attempt time-saving shortcuts in solving problems.

A caveat should be inserted here. Gender-linked variables do not necessarily translate into gender-linked effects on the critical outcomes. For example, the only causal relationship Sherman and Fennema (1977) could find among math usefulness, performance, and course plans was a moderate one between math being perceived as a male domain and math achievement, and the relationship was only for girls in the high school years.

This caveat notwithstanding, by the end of high school, these perceptual differences are associated with gender disparities in decisions made related to science and engineering. Perceptions of math usefulness generally predict intentions to take math courses and math achievement although the greater course-taking propensity of males is not clearly detectable until grade 12 and is not strong in predicting enrollment in mathematics courses even at the collegiate undergraduate level (Sherman and Fennema, 1977; Meece et al., 1982).

That female self-efficacy is involved in these disparate changes, is of little doubt. Girls' self-esteem drops precipitously after adolescence, with the drop for Latinas being the greatest, while African-American girls do not decline in self-esteem (they do dislike schools and teachers) (AAUW, 1991; Sadker and Sadker, 1994).[5]

The manifestations of gender difference take several forms. Older boys rate their ability as the "secret" of their success in science and engineering whereas similarly situated girls cite their personal diligence, skill, and effort (e.g., Jones and Wheatley, 1990; Mans et al., 1989). If girls experience problems in math, they perceive this as "personal failure" (AAUW, 1991, p.13). Especially boys, but

[5]There may be an important clue here to understanding gender differences if it can be shown that the socialization of African-American girls differs from that of other girls and that African-American female socialization is similar to that of majority males.

girls as well, believe that male understanding of science and math is superior, even when it is not (Linn and Hyde, 1989; Orenstein, 1994). Boys overestimate their abilities; girls *under*estimate theirs.

The connection between self-concept or self-esteem and ability in science and mathematics is important. Students are more likely to enroll in optional math courses when they perceive themselves to possess high math ability or feel confident in math (Meece et al., 1982; Sherman, 1980) although relatively few studies attempt to determine whether the links are causal and when they do the results are far from clear (Meece et al., 1982).

Math and science come to be viewed as a male achievement domain (Meece et al., 1982; Ernest, 1976); the same is true for science (Mason, Kahle, and Gardner, 1991). [6] Boys consider math to be a masculine subject (Brush, 1980); however, girls are less likely to gender-type math than are boys and although girls may perceive math-related careers as masculine, they do not necessarily view math as inappropriate for them (Meece et al., 1982). They do believe that studying science is more important for boys and that it will be of greater utility for boys (Linn and Hyde, 1989).

But, again, these differences rarely are detected until the self-concept hierarchy years and they are not clearly evident until late adolescence. For example, gender-linked mathematics enrollment patterns materialize at grade 12 and are still weak among college undergraduates (Meece et al., 1982).

Clearly, female self-efficacy in science and math, on average, is less than for males. The explanations would appear to involve important gender-linked perceptual differences that take many forms but that usually involve perceived male-female role disparities.

The important policy question is, What changes between preadolescence and adolescence? Of course, differences in socialization experience are thought to explain many if not most of the gender differences, but what are the specific mechanisms that lead to gender differentiation in science and engineering?

Identification of specific causal agents is not far advanced. Attitudes of parents and other family members, teachers, counselors, and peers are variously known or believed to play some part but most evidence is more impressionistic than hard (Haven, 1971; Jacklin, 1979; Nash, 1979).

In regard to race, it appears that self-efficacy and self-confidence are primary factors in African-American male consideration of math/science careers whereas for females personal interests are paramount; nevertheless, interest levels are approximately equal (Post, Stewart, and Smith, 1991). Males give broad consideration to career possibilities, relative to females. Disturbingly, African Americans do not appear to reflect on their personal abilities in these considerations.

---

[6]Many of the studies that compose the core of existing research and present thinking are now quite dated. It is almost certain to be true that many gender-linked differences have changed as the women's movement has grown.}

The CIRP Results. What do our results say about these issues? The best measure of science, math, and engineering self-efficacy available in the CIRP files is at the point of college matriculation, when, as our findings show, the perception that one's preparation in math and natural science is "better than most" is associated about equally with the greater likelihood of selecting physical science/engineering or biological sciences, compared to other fields, as the first choice of college major. The marginal effect, that is the effect of holding all other variables in the analysis constant, is a substantial 1.6 percentage point increase in this probability for each 10 percentage point increase in the probability of having this perception. This finding is for all matriculants.

What about selected subgroups? First, the fact that white men are most likely to perceive their math and natural science preparation to be better than most is noteworthy. The comparative mean values are white men, 43%; white women, 31%; black men, 26%; black women, 16%; Hispanic men 23%, Hispanic women, 17%. These are substantial differences. It is little wonder that the numbers of women and minorities entering science and engineering are relatively small.

How do the effects of self perception vary with race and gender? The results reported in Tables 2-7 can be used to answer this question. These tables report the effects of the explanatory variables on the probabilities of selecting, as a first choice of college major, respectively, the physical sciences/engineering, biological sciences, liberal arts, business, and the residual category. The positive marginal effect of the science and math preparation variable on the probability of choosing a major in physical sciences/engineering is larger for white males than for white females, blacks, and Hispanic females. Compared with white males, only Hispanic males show a greater positive effect of perceived preparation in math and science on the probability of selecting physical sciences or engineering as the first choice major.

Turning to the biological sciences, first our stated expectations in regard to math and science preparation are met: Matriculation is somewhat less strongly associated with self-perception of above average math and science preparation, than is the case for the physical sciences and engineering matriculation. Further, we can see that a reported "stronger than most" math and science preparation has a larger positive effect on the probability of selecting biological sciences for white women than for white men. The same is true for black females versus black males. Hispanic males and Hispanic females are about equal in this regard.

It is clear that math and science (self-perceived) preparation is less important in selecting a biological sciences major than is the case for the physical sciences and engineering. Nevertheless, if preparation in math and science positively influences the probability of selecting a first major in the science fields in general, perceived better preparation in math and science must simultaneously reduce the probability of selecting a non science major. The largest negative effects were generally for the probability of selecting a major in liberal arts, especially for Hispanic males. In the case of Hispanic females, perceived preparation in math and science had very little effect on any of the probabilities of selecting a college major.

The results for the Hispanic men bear closer examination. This is a particularly interesting group for examining these questions because Hispanic values and socialization sharply differentiate Hispanic men from Hispanic women; that is, gender roles tend to be quite distinct among Hispanics. Traits held to be more masculine tend to be more clearly associated with males than is the case for whites, and those held to be more feminine tend to be more clearly associated with females. The "macho" male construct is widely known and appears to be more than stereotype, although the trait almost certainly is more characteristic of working class than non-working class Hispanic males, who admittedly are likely to be present in greater numbers among college-goers. This notwithstanding, considering all six groups, we would expect that self-concept in science and math would be most powerful in predicting male Hispanic matriculation patterns, and this is what occurs. Not only is this math and science perception variable most strongly associated with physical sciences and engineering matriculation for Hispanic males, the *absence* of such perception is most strongly associated with matriculation in the liberal arts. Considering both sides of the results for this variable, the weight of evidence is that math and science (self-perceived) preparation is related to whether physical sciences and engineering majors will be selected, particularly vis-à-vis the liberal arts. Further, the results for Hispanic males strongly reinforce the general importance of math and science self-efficacy.

There is one other variable in the data set that should impact directly on self-concept, if not self-efficacy. That variable is one's grades earned in college. We would expect that earning good college grades would be more important to white women than to white men because self-concept/self-efficacy is supposed to be more critical for the former group, than the latter, in the pursuit of science and engineering majors. Indeed, that is what the results show. For the physical sciences and engineering, the effect of having earned undergraduate grades of B or higher has almost twice the effect on achieving an undergraduate degree and almost a 50% greater effect on earning a masters degree for white women compared to white men. The patterns are of the same general order in the biological sciences; however, here, the effects of higher grades are more beneficial for white males in earning doctoral and professional degrees.

To what does this all sum? To the extent that self-perception of math and science preparation reflects math and science self-efficacy, there does appear to exist an association with first choice of college major, and the relationship to this self-efficacy is more clearly shown for the physical sciences and engineering than for the biological sciences. The effect is strongest for males; the order is Hispanic males, white males, and black males followed by the female groups as predicted. The differences, however, are small. The results for liberal arts majors provide additional insights. Low math and science self-efficacy (perceived preparation) is most sharply delineated between Hispanic males and Hispanic females in choosing liberal arts majors, a finding that we think points up the importance of self-efficacy because we would hypothesize the differing effects to be the largest for these two groups, due to their sharply different socialization experiences. Perceived better preparation in math and science has a larger discouraging effect on choosing liberal arts for black males than for black females. On the other hand, the reverse is true for whites: math and science preparation has a somewhat larger negative effect on the probability of choosing a major in liberal arts for white females than for white males.

Our sense is that the findings clearly support the importance of math and science self-efficacy in determining whether individuals will enter science careers, especially in the physical sciences, and that the findings also support the general outline of how females are affected differently from males, as maintained in the literature, although the strength of that support is fairly modest. Also, we have found support for the importance of this construct for minority students, a conclusion that previously has not been examined closely and frequently, according to our reading of the literature. Clearly, the "perceived preparation in math and science" variable is not a precise measure of math and science self-efficacy nor would we expect it to be a particularly powerful one. Further, the effects found are modest and are examined only for first choice of college major. We hypothesize that with more sensitive gauges of the concept and examination of the concept during the college years, the relationship of the importance of math and science self-efficacy to persistence in science and engineering would be consistent and stronger than our data show. This is an arguable point but one we subscribe to quite strongly after having been immersed for more than two-and-one-half years in the literature and the data.

Other information in the CIRP data set permit some pursuit of the self-efficacy concept but not specifically in science and engineering, and therefore is of limited relevance to our purpose. After choice of major, the next possible impact of self-efficacy is reflected in the data for what degrees, if any, students earn. Here, "Intellectual self-confidence" probably is a good measure of general academic self-concept, and the variable is associated, modestly, with earning degrees in science and engineering. But as we suspected, the variable does not differentiate at all well between white males and white females in science and engineering, which were the only sub-groups for which CIRP data permitted statistical contrasts. We would only expect such differentiations if the variable tested *mathematical or scientific* intellectual self-confidence. Numerous other variables in the data set might be of some relevance to one's self-efficacy in math and science but the connections almost always are less than direct and fail to differentiate between types of self-efficacies, i.e., mathematical and verbal.

1.2 Complementary Effects

Complementary effects of self-efficacy may work against the science and engineering achievements of women and perhaps of minorities: Ample evidence exists suggesting that individuals competent in science and math but nevertheless possessing weak math self-concepts may tend to compensate by choosing fields in which their stronger verbal self-concepts may prove more useful, in other words where they can be more successful. For example, in their comprehensive review of the self-efficacy literature on career selection, Lent and Hackett (1987) cited "gender by task interactions" that ran counter to predictions, of task failure in one domain being linked positively to self-efficacy expectations in the other domain (p. 361). The question here is whether this conclusion pertains to gender or racial/ethnic differences in science and engineering outcomes in later years.

Our results did not lend support to the complementary effect. The CIRP files included a variable reflecting whether the individual perceived their pre-college preparation in reading, composition, and writing as "better than most." We tested this variable for both physical sciences/engineering majors and biological sciences majors, hypothesizing that the gender-linked effects would be more clearly pronounced in the former than in the latter, due to the presumed greater importance of math in the former and the known greater student criticism of physical sciences and engineering courses and programs. Though small, marginal effects were found for the biological sciences, possessing a relatively strong verbal (reading, composition, writing) self-concept was negatively associated with earning a biological sciences degree, as hypothesized; however, this result was due to the patterns for white males, not white females. In other words the "pull" of their verbal ability did not attract white females out of the biological sciences. Our interpretation for the biological sciences is that relatively weak math self-concepts are not a substantial problem for white females because these disciplines are not particularly dependent upon mathematics competence nor are the characteristics of the biological sciences particularly disagreeable to women. The lack of a clear pattern in the physical sciences was a greater surprise. Perhaps, for physical sciences or engineering majors, any complementary effects were reflected in other factors, in the analysis, that co-varied with verbal ability; or, perhaps individuals with high verbal self-concepts do not enter these fields in the first place. Of course, it may also be that the complementarity principle is not valid at all although we doubt that this is the case.

1.3. Family Influences.

Although our full report to NSF explores in detail numerous effects of family on science and engineering outcomes, we discuss here only those effects that appear clearly related to two of our three constructs: self-efficacy and, in turn, commitment. Most of the related research, which is of a correlational rather than causal nature, points to the importance of family traits, in particular, demographic traits such as education levels and occupations. For example, Jackson, Gardner, and Sullivan (1993) report that women who enter male-dominated fields such as science and engineering tend to come from intact families, have mothers who work, and have parents who are well educated and who consider success to be important. Worthley (1992), too, reports that science persistence is associated with having highly educated parents. The Jackson et al. work finds that women who become engineers are likely to have fathers who are engineers and they are likely to marry engineers. The AAUW study (1991) reports that families (and schools) are the greatest effectors of science and engineering aspirations, and Astin (1993), from a more recent CIRP file, finds a positive relationship between majoring in engineering and having a father who is an engineer. The engineering students in Astin's sample are clear that their parents want them to become engineers.

The CIRP Results. We single out one variable, parental occupations, for its likely connection to self-efficacy and commitment to science and engineering. We find that having a parent engaged in science and engineering occupations increases the probability of selecting a major in science and engineering, as a first choice. The differences among the six groups in these regards are particularly informative. The effects are uniformly greater for males than females, and are by far the largest for Hispanic males, followed by black males. A connection to self-efficacy is suggested, especially for the Hispanic males, again supporting our observations regarding Hispanic culture. This connection may work through another powerful force identified in the literature, that of commitment. We hypothesize that for minority men, especially Hispanics, having a parent who is employed in physical sciences or engineering is a powerful factor in the student's perception that such an occupation is a realistic goal, and that this perception reinforces science self-efficacy and supports the student becoming committed to that goal. For white men the marginal effects are not as great; the reason probably is that occupational identification and efficacy are not as vital to white males in selecting a major in these fields: White men are less likely than minority men to perceive limits on opportunities among fields of employment. It would be consistent with the literature to explain the quite different effects for women, minority and majority, as reflecting a more "eclectic" decision structure; that is, the fact that women tend to be more conflicted between career and family values (a topic expanded upon below) may reduce the significance of the career feasibility factor that is apparently important to minority men. Again, this is for the physical science and engineering. For the biological sciences the patterns are far more mixed for women although the magnitude of the effects often are similarly strong.

It is particularly enlightening and instructive to policy that the marginal effects of fathers' occupations in science and engineering on the science and engineering outcomes for their children are even larger later, in shaping employment patterns, than they are in selection of first majors. Clearly, individuals working in science and engineering occupations in 1980 were substantially more likely to have had fathers whose occupations were science and engineering than some other field. Further, the effects for the entire sample of white men and white women were almost totally a function of the effects for the males. For example, increasing by 10 percentage points the probability of having such fathers increased the probability of white males being employed in science and engineering occupations by almost 5 percentage points, but for white females the results were essentially zero. In a pattern that became clear when we examined all the "family influence" variables, white men appear more responsive to such variables than do white women.

1.4. Parental Degrees

One related finding concerns the disparate effects of parents' educational level on (the subjects') higher degree attainment in the physical science or engineering. This degree attainment was associated positively, for both white males and white females (clearly more so for males), with whether fathers were college graduates; however, the results were reversed for white females and white males for mother's college degree status—positive effects of mother's possessing a college degree for females and negative for males. In light of the literature, our interpretation of this finding is that since white women, on average, have weaker math/science self-concepts, compared to white men, the benefit of the mother as a role model might be more important for women involved in science and engineering than for similarly-situated men. Further, for those in the physical sciences or engineering, the effects of parental degrees persist at least through the masters degree; in fact, they grow modestly stronger for both men and women, before declining at the doctoral and professional level. The sociological literature emphasizes the importance of parental achievement in determining whether offspring will "see themselves" (self-efficacy) in some particular professional role and thus pursue that career.

For the biological sciences, the patterns are again more varied. For white men, achieving higher degree levels is associated positively with one's father holding a college degree, but the relationship is consistently negative for one's mother holding the degree. For women the patterns are very weak. This provides still more evidence that individuals who pursue study and work in the biological sciences are substantially different from those in the physical sciences and engineering.

2. Peers

Combined with physiological change, peer influence, we believe, is the "missing link" that explains the great transformation of girls, at adolescence, from being full and equal participants in science and math to becoming reluctant players. In our early reading of the literature, we concluded that peers were only one of a multitude of factors or variables playing some fairly modest part in shaping the science and engineering decisions of women and minorities. Then, a book by two anthropologists, Dorothy Holland and Margaret Eisenhart (*Educated in Romance,* 1990), came to our attention. The book greatly changed our views of the importance of peers, in particular for women. It made connections that we had puzzled over for some time. Specifically, Holland and Eisenhart helped us understand how female self-concept, self-efficacy, classroom experiences, and external goal orientation come together to deny women their representative places in the science and engineering professions.

Although the focus of the Holland and Eisenhart study is on the college years, the origins of the phenomena they observe are in the onset of adolescence, which marks not only a clear demarcation between boys and girls in regard to science- and math-related behaviors, it signals the beginning of distinctly different paths of gender-linked general personality development.

From its nationwide survey, the American Association of University Women (AAUW, 1991) concludes that adolescence is a critical time for female self-identity development. It is a time of both dramatic biological and psychological changes; it is a "critical moment" in the development of one's lifetime choices and decisions (AAUW, p.2). The AAUW survey shows that prior to adolescence, girls are "confident, assertive and feel authoritative about themselves" (p. 4). In subsequent years their self-esteem declines dramatically. Notably, adolescent girls rank "being popular" as the most important personal concern, whereas boys list competence and independence. Girls report not being happy "with the way I am" (p. 5), growing more timid and tentative, and becoming more conflicted. Their physical appearance is of great importance to them (AAUW, 1991).

The similarity of these observations about self, in general, to girls' self-efficacy in science and mathematics is striking. Lack of self-confidence, low self-esteem, timidity and tentativeness, are all descriptions that fit many post-adolescent girls in regard to science and mathematics, as is the perception on the part of girls that boys are more able in math and science than they are.

Logically, the dramatic decline in female self-esteem that occurs during adolescence (AAUW, 1991) would be expected to cause girls to become more other- rather than self-directed. Their primary concern with popularity suggests strongly that girls are likely to be particularly mindful of what their peers think of them: how they behave, their values and orientations, what they like and do not like. This other-directedness during adolescence might be expected to result in openness to the suggestions of parents and teachers, provided that those suggestions do not run counter to the values of the peer culture, in which case the outcomes might be far less certain. Nevertheless, generally, one might assume that this other-directedness influences girls to develop superior academic habits, for example to complete homework assignments, study dutifully, and generally be more diligent in academic matters. Clearly, the AAUW data suggest that girls may lack the self-confidence to set their goals independently, that they may pursue life goals reflective of what others, not they, want them to do. Again, the importance of commitment is clearly noted.

If adolescent girls are so susceptible to the influences of others, especially peers, what is it that others cause girls to do? Teachers, and to some degree parents, of course want girls to be good students, often even to study science and math. Female peers, on the other hand, according to the AAUW data, primarily appear to value popularity. Because this and related values emerge dramatically during adolescence, popularity almost certainly is in part a reflection of personal appearance and by implication, involves appeal or attractiveness to boys (Smith, 1992).

Holland and Eisenhart (1990) demonstrate that females gain much of their self-esteem through their relationships with males. The effects are circular: Validation of girls' and later women's behaviors by males leads to greater acceptance by peers, all tending to reinforce female perceptions that their sense of self-worth is connected importantly to the perceptions of males. This means conforming to what both female and male peers conceive of as appropriate female behaviors and values. On average, this suggests a lessening, not a strengthening, of girls' commitment to science and math; why, to a considerable degree, girls begin to lose interest in science and mathematics beginning in the junior high years; and why girls come to view science and mathematics as the domain of boys.

In discussions and presentations, our academic colleagues and our students initially expressed skepticism about this conclusions from this literature. They questioned the implicit assumption that girls were more concerned about and affected by boys than vice versa. Certainly, adolescent as well as post-adolescent boys, are vitally interested in girls. What differs, we believe, is how this interest plays out in relative commitment to science and engineering occupations, how males come to view work almost as an obligation because they see themselves as family providers, whereas girls often come to view work and family as equally attractive alternatives.

The CIRP Results. At the time of college entrance, CIRP collected information on one variable that may serve as a proxy for the degree to which peers might already have influenced individual's commitment to careers. The variable was whether the individual reported "some chance or a very good chance that he or she will marry while in college or within a year after college." This variable, of course, was an imperfect measure of peer influence: Although one could very easily be greatly influenced by peers in regard to the relative importance of relations with the opposite sex, one might still not plan or consider marriage during this time period; or, one could easily be influenced by peers and thus devote considerable time and effort to relations with the opposite sex without that influence necessarily manifesting itself in plans for or expectations of marriage. In short we anticipated modest effects from this measure, and that is what we found.

The small, negative coefficient for the "marriage variable" for biological sciences but small, positive coefficient for the other fields of first major, suggests that students in the biological sciences are relatively less interested in marriage and thus more committed to careers. Of course, the real testing of the variable comes in examining male/female differences. Here, the clearest results were for white females with respect to liberal arts and white males with respect to business. Among white females, having plans for marriage in college has the largest positive effect on the probability of selecting liberal arts with much smaller positive effects on the probabilities of selecting physical sciences/engineering or the residual category. Among white males, the only positive marginal effect of having plans for marriage in college is on the probability of selecting business; the marriage variable had a negative effect on the probability of selecting each of the other majors, including science and engineering. By contrast the marriage variable had a small positive effect on the probability of selecting physical sciences/engineering for white females. These findings are consistent with the notion of less commitment to physical sciences/engineering careers among white women. Black women with expectations of marrying while in college were especially less likely to declare biological sciences as their first choice of major. Among Hispanic men marriage had a positive effect on the probability of selecting science and engineering, whereas for Hispanic women marriage plans exhibited the largest negative effect on the probability of selecting biological sciences and exhibited the largest positive effect on the probability of choosing "Other," which would include the "undecided major" category and might indicate a relative lack of commitment to earning a degree. Our question here is how college peers might impact those plans.

In the 1980 CIRP follow-up survey, respondents were asked whether they had ever married. Arguably, this measure probably provides a better test of possible peer influence on women, in college and before, although to be sure there are many factors other than peer influence that might affect decisions to marry. In any case the results *for the total sample* are strongly in support of the peer influence hypothesis: Across the board, marriage impacted the probability of obtaining a college degree far more for women than for men. Among whites the impact was ten times as large for women; among black women it was twice as large; and, consistent with observations made elsewhere herein, Hispanic women were far less likely to graduate if married while Hispanic men were the only group to show a negative effect from having never married. In other words having married increased substantially the likelihood that Hispanic men would have achieved a degree. These results are strong support, not only for the peer influence variable, but also for the role of culture—if indeed these influences can be separated at all.

What about the science and engineering students? If our reasoning is correct, peer influence should be fairly powerful in inducing women to switch from science and engineering, at least from the physical sciences and engineering. The original survey in 1971 contained some highly relevant information related to the peer influence hypothesis, which holds that in placing a high value on the importance of being attractive to and devoting major efforts to relations with males, female peers impact negatively the career commitment of women to science and engineering. The variables in question reflected incoming students' self-ratings in regard to general popularity, popularity with the opposite sex, intellectual self-confidence, social self-confidence, and sensitivity to criticism.

Although the effects were small, the results appeared to be consistent with the hypothesized effects of peer influences upon degree attainment in physical sciences and engineering among white women. Among the five variables, popularity with the opposite sex was clearly the most potent predictor of *not* achieving degrees in the physical sciences and engineering; general popularity also was a negative factor. Our interpretation is that white females who perceive of themselves as popular are most susceptible to peer influences. Consistently, social self-confidence was a *positive* predictor of earning degrees in physical sciences and engineering, but interestingly, *intellectual* self-confidence was of relative unimportance. Our favored explanation for the former finding is that socially self-confident women, in contrast with those who self-describe themselves as "popular," are less susceptible to peer influences and as a result are more committed to career goals. The fact that these patterns did not hold, but rather were very weak for white women in the biological sciences, was consistent with observations made elsewhere herein regarding why these women were less "at risk" than women in physical sciences and engineering. Finally, the modest coefficients for these variables among women in science and engineering almost certainly reflect major self-selection effects; that is, those who place a high value on attractiveness to men are not likely to enter science and engineering in the first place.

Continuing, having never been married was only very slightly associated with earning degrees in science and engineering for the entire sample. But, as expected, having married did positively affect earning degrees in physical sciences and engineering among white men and negatively affect such achievements among white women. Put another way, at the margin, these women "pay a cost" for getting married; men in physical sciences and engineering reap a dividend. We think that many of these men are supported both financially and morally by their wives and further that many are motivated to succeed by anticipating that they will be supporting families after graduation. Similarly-situated white women, we think, are proportionately more likely to forego or defer their own career plans to these same ends. Contrary to our expectations, the marginal effects among white women were very slightly larger in the biological sciences than in the physical sciences and engineering.

Some additional insights as to what may be operating may be gained by examining yet another related variable: parenthood. The CIRP files contain information as to the sample member's "number of children" in 1980. First of all, having children had an overall (that is, the overall sample) negative impact on degree attainment in both the physical sciences and engineering and in the biological sciences. This result may suggest, at least partially, an economic explanation, i.e., the cost of child rearing and of opportunity costs: the cost of continuing in and achieving higher degrees in science and engineering. This reasoning is supported by the finding that the effects of parenthood in the physical sciences and engineering are negative for both white men and white women and that the magnitude of the effects are larger for white women the more children they have. If there are three or more children, the effect is profound. The results for the biological sciences reinforce the economic rationale although the results are not as strong.

The CIRP data also permit pursuit of the effect of parenthood upon employment in science and engineering. Overall, having children reduces the probability of working in science and engineering and the effects are largest if three or more children are involved, especially for white women. For men, overall, parenthood may *enhance* the likelihood of working in science and engineering; the explanation probably is again largely, perceived economic necessity.

Finally, we may examine labor force status of science and engineering majors. Overall, employment status essentially is unrelated to having been married. Having children is associated negatively with full-time work in science and engineering and positively with other work statuses, suggesting that child rearing takes individuals out of the science and engineering labor force. The question is who is affected and in what direction? Having never been married reduces the probability of full-time employment in science and engineering and increases the probability for the other categories (part-time employment, unemployment, and out of labor force) for white males, but increases the probability of the former and decreases the probability of the latter for white females. These patterns, too, may argue for economic explanations. Lack of family responsibilities perhaps allow white males greater economic freedom. For white women, such lack of responsibilities may permit women to work full-time because they do not have such responsibilities. There may be many other explanations.

Where does this leave us in regards to what we consider to be a powerful explanation for lower science and engineering participation for white women versus white men: the influence of peers? Clearly, the degree attainment and labor force participation variables are at best indirect measures of possible peer influence. There are many alternative explanations; economic factors, for example, clearly are involved to some degree. Nevertheless, these measures may be the ultimate test of the influence of peers in leading white women toward what Holland and Eisenhart (1990) call being *Educated in Romance* and away from science and engineering. Ultimately, these measures are what the study is all about: What leads women (and minorities) toward or away from earning degrees and working in science and engineering? At the very least, we may say that the findings do not permit rejection of the peer influence hypothesis. Although we are not fully satisfied with the measures available in the CIRP files, and although our methods are not as sensitive to peer effects as some of those described in the literature—e.g., the ethnographic methods of Holland and Eisenhart (1990)—the combination of our results and those from more sensitive methods leave us subscribing strongly to the peer influence hypothesis as a central part of the explanation for why many white females abandon, reduce aspirations toward, or never enter science and engineering. Our results for the impacts of marriage and child rearing add another important dimension to the previous work of other researchers.

3. Goal Commitment

Necessarily, much already has been said about goal commitment. Obviously, it is difficult to separate the effects of self-efficacy from peer influence and from commitment to personal goals. From the literature and from our findings, we infer that science and math self-efficacy form in elementary school but are retarded for women by the interaction of physiological changes and peer influences, resulting in reduced or conflicted commitment to science and engineering study and work.

Commitment is known to be the most potent predictor of persistence in almost all human endeavors (e.g., Ethington, Smart, and Pascarella, 1988; Sarkar, 1993). For women their interactions with parents, K-12 science and math teachers, and peers appear often lead to formulation of goals that are "externally" directed; that is, the important life goals of women frequently are formed largely in response to the desires of others. Women are socialized to seek to please. Thus it is not surprising that women tend to be interested in careers emphasizing human interaction, even if within science and engineering. The evidence is noteworthy.

Both men and women tend to view math and science more as "masculine" fields of study. For men the perception probably reinforces the tendency to persist in these fields; for women the reverse probably is true. Women may be attracted to fields seen as more nurturing, fields such as the social sciences and humanities (Hackett and Betz, 1981). Men exhibit nurturing behaviors too, but predominantly when that nurturing is in fields perceived to be masculine. For example, men are quite willing to be helpful and supportive of women in math and science classrooms and on math and science homework.

From her study with Hewitt, Seymour (1992) concludes that female self-worth in science and engineering is, in fact, extrinsically based, and that for many females, selection of a college major in science and engineering is externally rather than internally driven. Persistence, too, is connected directly to degree of goal commitment, specifically in engineering (Jackson, Gardner, and Sullivan, 1993). Thus, women, more so than men, require a "genuine interest" in their chosen careers (Dick and Rallis, 1991). The conclusion that is relevant to this discussion is that for many women commitment to science and engineering goals may be quite tenuous.

Men are shown to be more "single minded" than women in regard to work and careers (Eccles, 1987; Manis et al., 1989), a characteristic suggesting greater science and engineering goal commitment. *Women perceive an incompatibility between careers in science and family life, and they see raising a family as an attractive alternative* (Ware, 1988). The effects may be direct. Marion and Coladarci (1993) demonstrate that those women who place a high future value on family are less likely to take science and engineering courses; Seymour (1992) finds that female science and engineering students switch majors due to lack of personal goal commitment; and Lewis (1991) reports that those women who do wind up studying mathematics are "extremely job oriented" (p. 722). Interestingly and in some contrast, from a national sample Rowe (1993) find that neither family orientation nor perceived intrinsic value of jobs are particularly strong factors in predicting female selection of an engineering major.

The literature demonstrates that generally found male/female differences regarding career and commitment continue into employment. Employed female scientists and engineers are less involved with their work than are their male counterparts although the gap narrows for those holding master's and doctoral degrees; for both bachelor's and master's degree holders, time for personal lives is rated as more important by women, compared to men (DiTomaso and Farris, 1994).

Tobias (1992) also raises questions about the lack of role models in science and engineering and faculty prejudices against women, in particular bias against women who have strong family commitments. Although our search of the literature revealed some empirical support for this view, the amount of such evidence was less than we had expected; in some cases the *absence* of such problems in women's lists of grievances about science and engineering was conspicuous; and *causal* connections were not found. Most related information was only conjectural or anecdotal.

We would expect that any individual who is marginally committed to any goal, in this case a science and engineering degree study program, will be less likely to tolerate unsatisfactory or adverse conditions than one who is fully committed. In short, all else equal and on average, men should be more likely than women to persist in science and engineering.

The CIRP Results. We have noted already that one of the largest family influence on science and engineering outcomes was whether parents were engaged in science and engineering occupations, and the quite different effects for women versus men. We spoke of the more "eclectic" decision structures of women in these regards. Also, we have observed that female commitments to science and engineering goals may be reinforced by having a well-educated mother.

A more direct measure of commitment, available in the CIRP files, is one's stated purpose in attending college. Two related choices, a strong desire to gain a general education and to prepare for graduate or professional school, are the measures. Among those who had been enrolled in degree programs in the physical sciences and engineering, white women were less likely than white males to have expressed these goals at the time of admission. While the stated desire to attend graduate or professional school had positive effects on the probabilities of being in higher educational categories in physical sciences/engineering for both white males and white females, these effects were generally higher for white females. To us, these results emphasize the importance, to white females, of being committed to science and engineering, if they are to achieve these ends: At the time of admission, white women are less committed to science and engineering, generally, than are white men, but those white women who are committed, as attested to by their graduate or professional school goals, will achieve in science and engineering at relatively high levels.

Another direct measure of commitment is whether one specified science and engineering as their likely career at the time of entering college. In fact, specifying science and engineering as one's likely career in 1971 was strongly and positively associated with being employed in a science or engineering occupation in 1980. Further, the effect was noticeably larger for white males compared with white females. Our conclusion is that, on average, women who select science and engineering majors may be less committed to science and engineering careers, which is to say that they are more conflicted about career and family.

3.1. School Personnel and Techniques

There are several elements under this heading that may be seen as impacting both self-efficacy and goal commitment. Perhaps most important is the fact that teacher criticism and feedback may be greater for boys than for girls (Dweck et al., 1978; Orenstein, 1994). Dweck et al. (1978) conclude that when boys are criticized by science and math teachers, the criticisms are predominantly for neatness, conduct and effort, whereas for girls the criticisms occur less frequently, and when faults are found, they are more commonly for academic performance. The research of Sadker and Sadker (1994) supports this perspective whereas other evidence is only partially supportive. For example, Heller and Parsons (1981) and Fennema (1982) conclude that although boys are criticized more than girls, neither gender receives much in the way of teacher disapproval. The AAUW study (Orenstein, 1994) appears to support the Dweck et al. view rather strongly. Indirectly, these differences in teacher criticism, by gender, may be of considerable significance for persistence later in science and engineering.

It is well known that teachers ask fewer questions of girls and engage them less in classroom discussions (e.g., Becker, 1981; Orenstein, 1994; Sikes, 1971). A conclusion reached from one important study is that girls only answer teachers when they absolutely know an answer, and sometimes not even then (Orenstein, 1994). Boys are more willing to "speak out" and to argue with their teachers when they think they are right (AAUW, 1991, p.6). Although the effect of these gender differences previously has not been shown empirically, "performance accomplishment" is the central concept of self-efficacy theory and gaining feedback is instrumental to one's personal evaluation in this regard. Self-efficacy in turn impacts commitment.

What do the CIRP data tell us about this issue? In fact, the CIRP results show that white women who had frequently argued with teachers in high school were more likely, than those who had not, to earn degrees and progress to higher degrees in biological sciences. This is a particularly interesting finding because such argumentation had not predicted selection of either biological or physical sciences/engineering as a first choice of college major for either white females or white males.

3.2. Higher Education Teaching Personnel and Administrators

Much of the criticism of science and engineering, which is particularly heavy among students in the introductory courses, is directed at specific faculty traits. The issue here is how negative student reactions may play out differently as a result of variations in science and engineering goal commitment, by gender.

In a study of Colorado science and engineering students by Seymour and Hewitt (Seymour, 1992), science and engineering faculty were characterized as being "too impersonal," as uncaring, unapproachable, and intimidating (p. 290). Tobias (1992) describes scientists as being insensitive to women, women's isolation in science and engineering, and even the need to do anything about such problems. She believes the reason for these behaviors lies in the fact that the scientists themselves are strongly committed to science, enjoy the rigorous competition that characterizes science, and because they are successful scientists, hold that anyone planning a science career should have the same outlook. The contrast between university science and engineering faculty attitudes and those of K-12 teachers, especially toward able girls, is stark and probably is important to lack of female science and engineering commitment and thus persistence. The fact that girls receive much personal attention in high school is believed to contribute to the extrinsic, rather than intrinsic, motivations of women; and the lack of that attention later, in college, is believed to be instrumental to lack of female science and engineering persistence in college (Seymour, 1992).

The CIRP Results. The CIRP files contained only one variable related to faculty in higher education: whether the student knew a faculty member or administrator during their collegiate career. In fact, the variable was connected to earning science and engineering degrees in physical sciences and engineering but essentially was not so related in biological sciences. Concerning physical sciences and engineering, the variable predicted earning the BA, and the predictive power roughly doubled for the MA. The effect was positive but smaller for the Ph.D. and was negative for earning no degree or some other degree. Clearly, this supports the importance of faculty/administrative interaction on some personal level. Further, by implication, not knowing professors or administrators led to transferring from physical sciences and engineering or dropping out; in fact, the largest effect was on earning a degree in some other field, i.e., switching majors. Again, and most instructive, the results are clearest for white women, compared to white men. The effects of the variable are about four times as large for the former, relative to the latter, at both the BA and MA levels and are about twice as large at the Doctoral/Professional degree level.

In general men and women differ in their reaction to science and engineering study, independent of whether they persist (Seymour, 1992). This gender difference is consistent with our central theses. As a result of low grades, impersonal treatment by faculty (rather than faulty faculty pedagogical practices), and the "lifestyle associated with the science and engineering careers," women complain most about discouragement or loss of self-esteem; women complain secondarily, but more than men, about work overload, rapid pace of coursework, competitiveness, and poor extrinsic rewards associated with science and engineering careers (Seymour, 1992). Women, too, are more critical of science and engineering teaching than are men. In a University of Michigan study (Manis et al., 1989), women also were more likely than men to disapprove of the competitive science and engineering atmosphere and to complain about the grading system and the subject matter, which they found dry or boring.

The ultimate question in the Seymour and Hewitt study was whether there were gender differences among switchers from science and engineering. Here, a mixed patterns appeared (Seymour, 1992). Women were far more likely than men to list the discouragement/self-esteem factor as a reason for switching (78% v. 43%) and they were twice as likely as men to list as a factor conceptual (overall; the general approach) problems in science and engineering study. Men were slightly more likely than women to specify poor teaching as a cause for switching and they were more likely to list large class size as a factor. The incidence of men who specified the desire to switch careers was more than twice as high as the rate for women, and the male incidence of specifying "inadequate rewards" in light of the effort involved was four times as great.

Clearly, these patterns are complex and the available information is sketchy. Simply concluding that women are more conflicted about career versus family goals or that men are more career focused may be an oversimplification. Men may be more focused on a career but not necessarily on a specific career; thus, they may be quite willing to switch from a career in science and engineering to some other vocation in response to some particular science and engineering environmental condition. The generalization that seems to make the most sense to us is that men are strongly career focused albeit not necessarily to a particular career whereas although women are more conflicted about career versus family options, if they are committed to a career, they may be more persistent than men in pursuing that career.

For minorities the empirical evidence regarding the impact of pedagogical or curricular science and engineering features is very limited. A ubiquitous criticism in the literature is that most general retention programs (not exclusive to science and engineering) make the ethnocentric assumption that solutions for minorities will be the same as or at least similar to those for whites (Hartman et al., 1991).

An example concerns cooperative learning. On the one hand it is noted that under-represented minority students, in general, tend to work alone (Culotta, 1992). Thus, whereas Chinese students often participate in informal study groups, African-American students prefer to study alone (Treisman, 1985). On the other hand, there is some evidence that cooperative learning strategies can be successful with minority students. Writing for a special issue of *Science*, Selvin (1992) reflects on programs that have been successful in advancing minority students in (K-12) mathematics and, contrary to the findings of Treisman (1985), he identifies as the "common themes" of such programs, the connection of math principles to daily life, especially in the early grades, and emphasis upon cooperative learning, students working in teams in a friendly environment. Perhaps the seeming contradiction is explained by the fact that in most institutions minority students work alone in science and engineering precisely because they are minorities in numbers. Culotta (1992), in writing about the success in science and engineering of historically black colleges and universities (HBCUs) where African Americans are in the majority, notes that "when blacks—or any other group—are in the minority in school, they tend to work in isolation."

What is suggested in the way of solutions to pedagogical dysfunctions in science and engineering? The empirical findings of Astin and Astin (1992) suggest the promise of science and engineering undergraduate students conducting independent research, assisting faculty in their research, and assisting faculty in teaching. They also cite the value of faculty using active teaching/learning forms, interdisciplinary work, team teaching and incorporation of women's and ethnic perspectives in the general education component of the curriculum.

DISCUSSION

The findings reinforce what becomes a clear pattern of results. In the physical sciences and engineering, students in general are at some risk. The patterns are clearest for white females—perhaps largely because the data are inadequate to test fully the effects for minorities—and numerous interventions can be effective in reducing that risk for white females. However, students in the biological sciences may be at no greater risk than students in general, or at least interventions appear to be of relatively modest import, even for white females. In short most of the problems are in the physical sciences and engineering, not in the biological sciences.

In reviewing the science and engineering literature, we soon learned that success in identifying student problems, even if gender-linked, did not necessarily translate into transfer out of science and engineering. Indeed, Seymour (1992) reported that groups of Colorado science and engineering "switchers" and "non-switchers" whom she and Hewitt interviewed in focus groups were quite similar.

Both of these categories of students cited poor teaching, a too-rapid pace of the curriculum, and heavy course requirements as science and engineering problem areas; however, whereas 89% of all switchers complained of poor teaching in science and engineering, only about 30% indicated that such teaching had contributed directly to their departing these fields. The list of criticisms about teaching was long and sobering but at times surprising. Class size was seen as a very serious problem in first- and second-year courses, but was not listed as a cause for switching, nor were laboratories and the use of foreign teaching assistants. The length of study required to earn science and engineering degrees and competitive grading that was "destructive of self-esteem" (read: self-efficacy) were problems that were largest in engineering, but again, presence of these traits did not necessarily mean student transfer or "switching."

The importance of self-efficacy on career choices is well founded in the literature. Our own analysis reveals the role played by self-perception of the adequacy of one's background in math and science. Males are more likely than females to report being better prepared than most in math and science. Perhaps more importantly, self-perception of math and science preparation has a larger positive influence on the probability of choosing a major in the physical sciences/engineering for males than for females. On the other hand this variable has a larger positive influence on the probability of selecting a major in the biological sciences for females than for males. If it is desirable to encourage a shift of females toward the sciences, especially physical sciences and engineering, then women's self-perception of math and science preparation must be improved.

Self-perception is a function both of one's knowledge of math and science concepts and one's confidence in one's math and science preparation. Increasing both of these is essential to increasing self-perceptions in science and math. Special efforts to expose female and minority students to elective math and science courses in their pre college years is important to enhancing both the skill acquisition and confidence necessary to making science a feasible choice for a college major. To reap the benefits from skill acquisition, female and minority students must come to believe that they can effectively use math and science tools. We present below some ideas for the kind of environments that might contribute to improved science and engineering self-efficacy and commitment. Our focus is on white women, as mandated by the limits of our results.

Families clearly can be highly instrumental to the science and engineering related aspirations of their children. Self-efficacy is the articulating concept between the two. The literature contains much anecdotal and some systematic evidence showing that having a parent who is a scientist/engineer is an important predictor of a female or minority person pursing such a career. The CIRP data support this conclusion. Since female and minority persistence and career choices in science and engineering education are impacted by parental education and career choices, in the short run these linkages can be exploited by encouraging efforts to introduce offspring to their parent's work environments in the science and engineering areas and by targeting such youth for recruitment to science and engineering careers. In the longer term as more women and minorities enter science and engineering fields, the probabilities that their children will also chose science and engineering careers should increase.

The importance of peer influence to education and career choices is well documented in the literature. Our own analysis supports the general thrust of the literature through the estimated effects of marital plans, marital status, perceptions of social popularity, and numbers of children on the probabilities of choice of major, persistence to higher degrees in science and engineering, entry into science and engineering careers, and labor force status. An implication of all of these findings in relation to peer influences is that we must develop more and better interventions, especially support systems, if we are to increase the likelihood of women entering science and engineering. For example, it is not uncommon for some colleges and universities to structure student living arrangements according to various "themes". Careful consideration should be given to structuring dormitory arrangements so that female and minority science and engineering majors can be housed in close proximity to one another.

One of the clearest implications of extant research for educational policy purposes, however, is for earlier intervention. The fact that the science and math paths of girls begin to depart sharply from those of boys during the junior-high or middle-school years and that peer influences are paramount in explaining this departure argue strongly for powerful policy interventions. Somehow, peer pressures that turn girls' attention toward popularity, especially toward attractiveness to and acceptance by boys, largely as mediated by their female peers, must be addressed. Others, such as social psychologists and school counselors, are better equipped than we to identify the promising strategies; however, it is clear that early (preschool) socialization processes must be addressed and that strategies targeted on the adolescent years must be strongly proactive. We conjecture that the first step, in the adolescent years, must be to make young girls aware of how highly susceptible they are to the influences of their peers and how peers exercise their influences, that these influences may shape the critical decisions of their entire lives, particularly the career and family choices they make. What is clearly beyond our capabilities is to identify the interventions necessary to reinforce academic values, particularly those related to science and math, and to inhibit the negative effects of peers.

A number of policy suggestions for somewhat later in life are related to the Complementary Effect, described in the empirical literature. Although our findings failed to confirm this effect, we subscribe to the conclusion that even individuals highly able in science and engineering will gravitate to other fields in which they also have good aptitudes when difficulties in or disenchantment with the former are encountered. We believe that in the high school years, though well-meaning, science and math teachers fail to challenge young women as much as they should. Treating women as "the weaker sex" is no favor to them, at least when science and engineering careers are considered. Science and math teachers should interact with their female students with no more delicacy than they do their male students. They should challenge them, engage them rigorously, expect high performance of them. They should insist that females do their own lab work, answer difficult questions, and justify their responses; in short they should give their female students much more experience in the give-and-take of intellectual discourse and much less in the way of deferential, less-challenging treatment. There is nothing unfeminine about questioning or even arguing with teachers about science and math problems and issues.

If these approaches were fully carried out in pre college years, the problems of the college years would be more manageable. However, there still would remain the need to address many structural issues in college-level science and engineering programs, especially in the introductory courses. Here, too, our expertise is somewhat limited, or is at least largely dated. The fundamental issue for science and engineering education is how to make vital changes while maintaining high quality. Science and engineering course work is difficult. On average it does require more time and effort than does course work in most other fields. Science and engineering programs usually do attract more highly able students than do most other fields, and on average they do produce relatively competent graduates. Yet, the criticisms of these programs by their students, white men as well as others, cannot be fully discounted as merely a desire for less work and easier grading. We are not expert enough to conclude which polices and practices should be changed although we do have some tentative thoughts that are based on our work. We wonder, for example, whether there are not better alternatives to the almost exclusive reliance upon mathematically-based problem solving utilized in most introductory physical sciences courses, whether a greater mix of strategies might not yield overall superior results. We wonder why cooperative learning strategies could not be encouraged, and why science and engineering faculty might not be persuaded to encourage, rather than discourage students, to persist in science and engineering programs. Our sense is that if science and engineering faculty merely became aware of the implications of many present practices they would be able to design adaptive strategies that would increase, not decrease, their programs' quality, overall. Our observations in outstanding science and engineering programs in private liberal arts colleges tell us that effective, adaptive strategies are clearly possible when science and engineering units become dependent upon student enrollment levels if they are to remain viable.

We note the promise of one particular technique for environmental enhancement of science and engineering units. We observe from our findings that coming to know a professor or administrator is especially important for women in terms of raising the probabilities of persisting to higher degrees in the physical sciences and engineering. This finding supports considerable anecdotal evidence from successful science and engineering professionals, especially members of minority groups, who emphasize that it was a particular faculty member who had been instrumental in their science and engineering persistence and success.

As was the case for our findings, our conclusions related to goal commitment are closely connected to our conclusions about self-efficacy and peer influences. That is, the problems of goal conflicts for women and minorities emanate from early socialization experiences as mediated by self-efficacy and peer influences. White women, in particular, are conflicted about career and family because their family members, peers, teachers, and others provide conflicting inputs to them. In comparison with white men, this results in a smaller likelihood that white women will possess strong commitments to science and engineering. The result clearly is that women will be more vulnerable to all sorts of distractions or alternatives, both as students in science and engineering programs, in science and engineering careers, and before. One obvious response is to make girls aware, early on, of how their goals and aspirations are formed; what the costs of adherence to peer norms are, for example the costs of avoiding math and science courses in high school; and how they come to gain their perceptions of their own science and math abilities, skills, attitudes, and of the comparative attributes of males.

In closing it is important to note that in this work we have set aside the value judgment as to whether women, or for that matter minorities, ought to pursue educational and career goals identical to those of majority men, either in general or in science and engineering in particular. Rather, our intent is to shed light on what factors can improve the prospects of underrepresented groups entering and succeeding in science and engineering fields, should they choose to do so. It is our hope that this research can contribute to ways of improving the opportunities for women and minorities to make choices that could lead to careers in science and engineering.

APPENDIX

Analytic Variables for the CIRP Analysis

Dependent Variables:

COLLGRAD	= 1	if highest degree held is a bachelor's degree or higher
	= 0	otherwise
EMPLOY	= 0	if the individual is out of the labor force (unemployed, not looking for work)
	= 1	if the individual is unemployed (looking for work)
	= 2	if the individual is employed part-time
	= 3	if the individual is employed full-time
INCOME	= 0	if the individual's income < \$15,000 (\$1980)
	= 1	if the individual's income ≥ \$15,000 and < \$20,000
	= 2	if the individual's income ≥ \$20,000 and < \$30,000
	= 3	if the individual's income ≥ \$30,000
MULMAJA	= 0	if first choice for college major is in physical sciences (including mathematics, statistics, computer science, environmental science) or engineering
	= 1	if first choice for college major is in the biological or life sciences
	= 2	if first choice for college major is in the liberal arts (social sciences, humanities, fine arts
	= 3	if first choice for college major is business
	= 4	if first choice for college major is in the residual category
MULSAME	= 0	if graduated with the same major category as the first choice for college major
	= 1	if graduated with a different major category from the first choice for college major
	= 2	if did not graduate (highest degree, if any, is less than a bachelors)
SEIHIGH	= 0	if a 1971 freshman major in physical sciences/engineering and did not graduate with a four year degree in any field
	= 1	if a 1971 freshman major in physical sciences/engineering and holds at least a bachelor's degree but did not attain any degree in physical sciences/engineering
	= 2	if currently enrolled in a bachelors degree program in the physical sciences/engineering or possesses a bachelors degree in physical sciences/engineering and not currently enrolled in a higher degree program in physical sciences/engineering
	= 3	if currently enrolled in a masters degree program in the physical sciences/engineering or possesses a masters degree in physical sciences/engineering and not currently enrolled in a higher degree program in physical sciences/engineering
	= 4	if currently enrolled in a doctoral or advanced professional degree program in the physical sciences/engineering or possesses

		a doctoral or advanced professional degree in physical sciences/engineering
SEIIHIGH	= 0	if a 1971 freshman major in biological/life sciences and did not graduate with a four year degree in any field
	= 1	if a 1971 freshman major in biological/life sciences and holds at least a bachelor's degree but did not attain any degree in biological/life sciences
	= 2	if currently enrolled in a bachelors degree program in the biological/life sciences or possesses a bachelors degree in biological/life sciences and not currently enrolled in a higher degree program in biological/life sciences
	= 3	if currently enrolled in a masters degree program in the biological/life sciences or possesses a masters degree in biological/life sciences and not currently enrolled in a higher degree program in biological/life sciences
	= 4	if currently enrolled in a doctoral or advanced professional degree program in the biological/life sciences or possesses a doctoral or advanced professional degree in biological/life sciences
OCCUP	= 0	if current occupation is in the sciences (physical and biological) or engineering
	= 1	if current occupation is in the medical professions
	= 2	if current occupation is in the non medical professions
	= 3	if current occupation is in the field of business
	= 4	if current occupation is in the residual category
	= 5	if current occupation is in the blue collar category

Explanatory Variables:

ACTHS10	= 1	if frequently discussed future with parents in the year preceding the 1971 survey
	= 0	otherwise
ACTHS12	= 1	if frequently argued with teacher in class in the year preceding the 1971 survey
	= 0	otherwise
ACTHS34	= 1	if frequently asked teacher for advice in the year preceding the 1971 survey
	= 0	otherwise
ACTHS35	= 1	if frequently had vocational counseling in the year preceding the 1971 survey
	= 0	otherwise
ADVPBASC	= 1	if the highest degree in 1980 is an advanced professional degree in a nonscience/engineering field and reports an undergraduate major in science(physical or biological)/engineering
	= 0	otherwise

ADVPGDII	= 1	if the highest degree in 1980 is an advanced professional degree in biological/life sciences
	= 0	otherwise
ADVPRO	= 1	if highest degree in 1980 was an advanced professional degree
	= 0	otherwise
AG	= 1	if employed in 1980 in agriculture or forestry
	= 0	otherwise
BAGRDSC	= 1	if highest degree in 1980 is a Bachelor's degree in a nonscience/engineering field and reports a graduate major in science (physical or biological)/engineering
	= 0	otherwise
BASCI	= 1	if the highest degree in 1980 is a Bachelor's in physical sciences/engineering
	= 0	otherwise
BASCII	= 1	if the highest degree in 1980 is a Bachelor's in biological/life sciences
	= 0	otherwise
BFEMALE	= 1	if black female
	= 0	otherwise
BMALE	= 1	if black male
	= 0	otherwise
CAR71I	= 1	if probable career in 1971 was reported to be in the sciences (physical and biological) or engineering
	= 0	otherwise
CAR71II	= 1	if probable career in 1971 was reported to be in the medical professions
	= 0	otherwise
CAR71III	= 1	if probable career in 1971 was reported to be in the non medical professions
	= 0	otherwise
CAR71IV	= 1	if probable career in 1971 was reported to be in the field of business
	= 0	otherwise
CAR71V	= 1	if probable career in 1971 was reported to be either blue collar or in a residual field
	= 0	otherwise
CIVGOV	= 1	if employed in 1980 in government, civilian
	= 0	otherwise

COLGRADE	= 1	if average undergraduate grade was a B or higher
	= 0	otherwise
CPREPSCH	= 1	if high school had a college prep program
	= 0	otherwise
DADCOLGD	= 1	if father was a college graduate
	= 0	otherwise
DOC	= 1	if highest degree in 1980 was a doctoral degree
	= 0	otherwise
DOCBAI	= 1	if highest degree in 1980 is a Doctorate in a nonscience/engineering field and reports an undergraduate major in physical sciences/engineering
	= 0	otherwise
DOCSCI	= 1	if the highest degree in 1980 is a Doctorate in physical sciences/engineering
	= 0	otherwise
DOCSCII	=	if the highest degree in 1980 is a Doctorate in biological/life sciences
	= 0	otherwise
ELSECEDU	= 1	if employed in 1980 in elementary or secondary education
	= 0	otherwise
FIRE	= 1	if employed in 1980 in finance, insurance, or real estate
	= 0	otherwise
FULLTIME	= 1	if the respondent reports full-time employment in 1980
	= 0	otherwise
FUTFR10	= 1	if reported in the 1971 survey some chance or very good chance that he/she would join a social fraternity, sorority, club
	= 0	otherwise
FUTR1516	= 1	if reported in the 1971 survey some chance or very good chance that he/she would seek vocational counseling or individual counseling
	= 0	otherwise
FUTREMAR	= 1	if in 1971 the respondent reports that there is some chance or a very good chance that he or she will marry while in college or within a year after college
	= 0	otherwise
GENED	= 1	if respondent reported in 1971 that gaining a general education is a very important reason for his/her attending college
	= 0	otherwise

GRADPROF	= 1	if respondent reported in 1971 that preparing for graduate or professional school is a very important reason for his/her attending college
	= 0	otherwise
GRANTS	= 1	if respondent reported in 1971 that scholarships or grants would be a major source of his/her college financing
	= 0	otherwise
GRDSCI	= 1	if the highest degree in 1980 is a Master's or Doctorate in physical sciences/engineering
	= 0	otherwise
GRDSCII	= 1	if the highest degree in 1980 is a Master's or Doctorate in biological/life sciences
	= 0	otherwise
HFEMALE	= 1	if hispanic female
	= 0	otherwise
HIGHED	= 1	if employed in 1980 in higher education
	= 0	otherwise
HMALE	= 1	if hispanic male
	= 0	otherwise
HSPRIVUS	= 1	if attended a private, U.S. high school
	= 0	otherwise
HSRANK4	= 1	if ranked in the top quarter of high school class
	= 0	otherwise
HSRANK3	= 1	if ranked in the second quarter of high school class
	= 0	otherwise
HUMSERV	= 1	if employed in 1980 in human services
	= 0	otherwise
JOBCOL	= 1	if it is reported that college prepared one for the job in 1980 fairly well, well, or very well
	= 0	otherwise
JOBMAJ	= 1	if job in 1980 is reported as somewhat or closely related to one's college major
	= 0	otherwise
KIDS1	= 1	if reported 1 child present in 1980
	= 0	otherwise
KIDS2	= 1	if reported 2 children present in 1980
	= 0	otherwise

KIDS3	= 1	if reported 3 or more children present in 1980
	= 0	otherwise
KNEWPRAD	= 1	if reported in 1980 survey that he/she knew a professor or administrator
	= 0	otherwise
LOANS	= 1	if respondent reported in 1971 that NDEA, federal, college loans, or other repayable loans would be a major source of his/her college financing
	= 0	otherwise
MA	= 1	if highest degree in 1980 was a Master's degree
	= 0	otherwise
MABAI	= 1	if highest degree in 1980 is a Master's degree in a nonscience/ engineering field and reports an undergraduate major in physical sciences/engineering
	= 0	otherwise
MABAII	= 1	if highest degree in 1980 is a Master's degree in a nonscience/ engineering field and reports an undergraduate major in biological /life sciences
	= 0	otherwise
MAJ71IA	= 1	if first choice for college major in in the physical sciences (including mathematics, statistics, computer science, environmental science) or engineering
	= 0	otherwise
MAJ71IIA	= 1	if first choice for college major is in the biological or life sciences
	= 0	otherwise
MAJ71IIIA	= 1	if first choice for college major is in the liberal arts (social sciences, humanities, fine arts
	= 0	otherwise
MAJ71IVA	= 1	if first choice for college major is business
	= 0	otherwise
MAJ71VA	= 1	if first choice for college major is in the residual category
	= 0	otherwise
MANUF	= 1	if employed in 1980 in manufacturing
	= 0	otherwise
MASCI	= 1	if the highest degree in 1980 is a Master's in physical sciences/ engineering
	= 0	otherwise
MASCII	= 1	if the highest degree in 1980 is a Master's in biological/life sciences
	= 0	otherwise

MILGOV	= 1	if employed in 1980 in military
	= 0	otherwise
MOMCOLGD	= 1	if mother was a college graduate
	= 0	otherwise
MONEYJOB	= 1	if respondent reported in 1971 that obtaining money or a better job is a very important reason for his/her attending college.
	= 0	otherwise
NEVMAR	= 1	if never married as of the 1980 survey
	= 0	otherwise
NOCAR71	= 1	if probable career in 1971 was not reported
	= 0	otherwise
NOSAT01	= 1	if not satisfied with career counseling at last undergraduate college attended
	= 0	otherwise
NOSAT02	= 1	if not satisfied with personal counseling at last undergraduate college attended
	= 0	otherwise
NOSAT03	= 1	if not satisfied with tutoring at last undergraduate college attended
	= 0	otherwise
NOSAT06	= 1	if not satisfied with financial aid services at last undergraduate college attended
	= 0	otherwise
NOTOTFAC	= 1	if total enrollment, student/faculty ratio not reported
	= 0	otherwise
NRSAT01	= 1	if response missing for career counseling at last undergraduate college attended
	= 0	otherwise
NRSAT02	= 1	if response missing for personal counseling at last undergraduate college attended
	= 0	otherwise
NRSAT03	= 1	if response missing for tutoring at last undergraduate college attended
	= 0	otherwise
NRSAT06	= 1	if response missing for financial aid services at last undergraduate college attended
	= 0	otherwise
NRCLGRDE	= 1	if average undergraduate grade was not reported
	= 0	otherwise

NRNEVMAR	= 1	if marital status not reported in the 1980 survey
	= 0	otherwise
NRPREPRD	= 1	if response missing in 1980 survey regarding self rating on preparation for college in reading, composition, and writing
	= 0	otherwise
NRSELECT	= 1	if the CIRP college selectivity score is not reported
	= 0	otherwise
OCCUPI	= 1	if current occupation is in the sciences (physical and biological) or engineering
	= 0	otherwise
OCCUPII	= 1	if 1980 occupation is in the medical professions
	= 0	otherwise
OCCUPIII	= 1	if 1980 occupation is in the non medical professions
	= 0	otherwise
OCCUPIV	= 1	if 1980 occupation is in the field of business
	= 0	otherwise
OCCUPV	= 1	if 1980 occupation is in the residual category
	= 0	otherwise
OCCUPVI	= 1	if 1980 occupation is in the blue collar category
	= 0	otherwise
OCUPIF	= 1	if father's occupation is in the sciences (physical and biological) or engineering
	= 0	otherwise
OCUPIIF	= 1	if father's occupation is in the medical professions
	= 0	otherwise
OCUPIIIF	= 1	if father's occupation is in the non medical professions
	= 0	otherwise
OCUPIVF	= 1	if father's occupation is in the field of business
	= 0	otherwise
OCUPVF	= 1	if father's occupation is in the residual category
	= 0	otherwise
OCUPVIF	= 1	if father's occupation is in the blue collar category
	= 0	otherwise
OCUPFNR	= 1	if father's occupation is not reported
	= 0	otherwise

OTHDEG	= 1	if the highest degree in 1980 is in a non science/engineering field
	= 0	otherwise
OTHERIND	= 1	if employed in 1980 in the residual category
	= 0	otherwise
PARENT	= 1	if respondent reported in 1971 that parental resources would be a major source of his/her college financing
	= 0	otherwise
PAROCSE	= 1	if either parent's occupation was in the science and engineering fields
	= 0	otherwise
PARTTIME	= 1	if the respondent reports part-time employment in 1980
	= 0	otherwise
PARINC1	= 1	if parental income in 1971 < \$4,000
	= 0	otherwise
PARINC2	= 1	if \$4,000 ≤ parental income in 1971 ≤ \$5,999
	= 0	otherwise
PARINC3	= 1	if \$6,000 ≤ parental income in 1971 ≤ \$7,999
	= 0	otherwise
PARINC4	= 1	if \$8,000 ≤ parental income in 1971 ≤ \$9,999
	= 0	otherwise
PARINC5	= 1	if \$10,000 ≤ parental income in 1971 ≤ \$12,499
	= 0	otherwise
PARINC6	= 1	if \$12,500 ≤ parental income in 1971 ≤ \$14,999
	= 0	otherwise
PARINC8	= 1	if \$20,000 ≤ parental income in 1971 ≤ \$24,999
	= 0	otherwise
PARINC9	= 1	if \$25,000 ≤ parental income in 1971 ≤ \$29,999
	= 0	otherwise
PARINC10	= 1	if \$30,000 ≤ parental income in 1971 ≤ \$34,999
	= 0	otherwise
PARINC11	= 1	if \$35,000 ≤ parental income in 1971 ≤ \$39,999
	= 0	otherwise
PARINC12	= 1	if \$40,000 ≤ parental income in 1971
	= 0	otherwise
PARINC9H	= 1	if \$25,000 ≤ parental income in 1971
	= 0	otherwise

PREPRDWR	= 1	if preparation for college in reading, composition, and writing was self reported as better than most
	= 0	otherwise
PREPSE	= 1	if preparation for college in math and natural science was self reported as better than most
	= 0	otherwise
RATEFR13	= 1	if rated self in the 1971 survey as above average or in the highest 10% on popularity
	= 0	otherwise
RATEFR14	= 1	if rated self in the 1971 survey as above average or in the highest 10% on popularity with opposite sex
	= 0	otherwise
RATEFR16	= 1	if rated self in the 1971 survey as above average or in the highest 10% on self confidence (intellectual)
	= 0	otherwise
RATEFR17	= 1	if rated self in the 1971 survey as above average or in the highest 10% on self confidence (social)
	= 0	otherwise
RATEFR18	= 1	if rated self in the 1971 survey as above average or in the highest 10% on sensitivity to criticism
	= 0	otherwise
SELECT1	= 1	if the CIRP college selectivity score $<$ 1,000
	= 0	otherwise
SELECT2	= 1	if 1,000 $\leq$ the CIRP college selectivity score $<$ 1,300
	= 0	otherwise
SELECT3	= 1	1,300 $\leq$ the CIRP college selectivity score
	= 0	otherwise
SFRATIO1	= 1	if student/faculty ratio at 1971 college $<$ 10:1
	= 0	otherwise
SFRATIO2	= 1	if 10:1 $\leq$ student/faculty ratio at 1971 college $\leq$ 12:1
	= 0	otherwise
SFRATIO3	= 1	if 13:1 $\leq$ student/faculty ratio at 1971 college $\leq$ 15:1
	= 0	otherwise
TOTENR1	= 1	if total enrollment of 1971 institution $\leq$ 1,999
	= 0	otherwise
TOTENR2	= 1	if 2,000 $\leq$ total enrollment of 1971 institution $\leq$ 4,999
	= 0	otherwise

TOTENR3	= 1	if 5,000 ≤ total enrollment of 1971 institution ≤ 9,999
	= 0	otherwise
TOTENR4	= 1	if 10,000 ≤ total enrollment of 1971 institution ≤ 19,999
	= 0	otherwise
TOTENR2A	= 1	if 1,000 ≤ total enrollment of 1971 institution ≤ 9,999
	= 0	otherwise
TOTENR3A	= 1	if 10,000 ≤ total enrollment of 1971 institution ≤ 19,999
	= 0	otherwise
TOTENR4A	= 1	if 20,000 ≤ total enrollment of 1971 institution
	= 0	otherwise
TRADE	= 1	if employed in 1980 in wholesale or retail trade
	= 0	otherwise
TRCOMUT	= 1	if employed in 1980 in transportation or public utilities
	= 0	otherwise
UNEMPF	= 1	if father is unemployed
	= 0	otherwise
WFEMALE	= 1	if white female
	= 0	otherwise
WORKSAV	= 1	if respondent reported in 1971 that work, savings, or GI benefits would be a major source of his/her college financing
	= 0	otherwise
WRKBA	= 1	if working on a Bachelor's degree in 1980
	= 0	otherwise
WRKMA	= 1	if working on a Master's degree in 1980
	= 0	otherwise
WRKDOC	= 1	if working on a Doctoral degree in 1980
	= 0	otherwise
WRKADVPR	= 1	if working on an advanced professional degree in 1980
	= 0	otherwise

References

American Association of University Women. (1991). *Shortchanging Girls, Shortchanging America: A Nationwide Poll to Assess Self-Esteem, Educational Experiences, Interest in Math and Science, and Career Aspirations of Girls and Boys Ages 9-15.* Washington, D.C.: AAUW.

Astin, A. W. (1993). *What matters in college?: Four critical years revisited.* San Francisco: Jossey-Bass.

Astin, A. W., and Astin, H. S. (1992). Undergraduate science education: The impact of different college environments on the educational pipeline in the sciences. *Final Report to the National Science Foundation (Grant No. SPA-8955365).* Los Angeles: The Higher Education Research Institute, UCLA.

Bandura, A. (1977). Self-efficacy theory in human agency. *American Psychologist* 37: 191-215.

Becker, J. R. (1981). Differential treatment of females and males in mathematics classes. *Journal for Research in Mathematics Education* 12: 40-53.

Benbow, C. P., and Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science* 222: 1029-1031.

Betz, N. E., and Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Behavior* 23: 329-345.

Brush, L. R. (1980). *Encouraging girls in mathematics: The problem and the solution.* Cambridge, MA: Abt Associates.

Castle, E. M. (1993). Minority student attrition research: Higher education's challenge for human resource development. *Educational Researcher* 22: 24-30.

Commission on Professionals in Science and Technology. (1994). *Professional Women and Minorities: A Total Human Resource Data Compendium.* Eleventh Edition. Washington, DC: Commission on Professionals in Science and Technology.

Culotta, E. (1992). Black colleges cultivate scientists. *Science* 258: 1216-1217.

Dick, T. P., and Rallis, S. F. (1991). Factors and influences on high school students' career choices. *Journal of Research in Mathematics Education* 22: 281-292.

DiTomaso, N., and Farris, G. F. (1994). *Degrees and diversity at work.* IEEE Spectrum, April, 38-41.

Dweck, C. S., Davidson, W., Nelson, S., and Enna, B. (1978). Sex differences in learned helplessness: II. The contingencies of evaluative feedback in the classroom and III. An experimental analysis. *Developmental Psychology* 14: 168-276.

Eccles, J. S. (1987). Gender roles and achievement patterns: An expectancy value perspective. In J. M. Reinish, L. A. Rosenblum, and S.A. Sanders (eds.), *Masculinity/femininity: Basic perspectives* (pp. 240-280). New York: Oxford University Press.

Ernest, J. (1976). *Mathematics and sex.* Santa Barbara, CA: University of California Press.

Ethington, C. A., Smart, J. C., and Pascarella, E. T. (1988). Influences on women's entry into male-dominated occupations. *Higher Education* 17: 545-562.

Ethington, C.A. (1995). Navigating the pipeline: Women's interpretations of socialization influences in science, engineering, and mathematics. Paper presented at the Association for the Study of Higher Education, Orlando, Florida, November, 1995.

Fennema, E. (1982). Attribution theory and achievement in mathematics. In S. R. Yussen (ed.), *The Development of Reflection.* New York: Academic Press.

Friedman, L. (1989). Mathematics and the gender gap: A meta-analysis of recent studies on sex differences in mathematical tasks. *Review of Educational Research* 59: 185-213.

Hackett, G., and Betz, N. E. (1981). A self-efficacy approach to the career development of women. *Journal of Vocational Behavior* 18: 326-339.

Hackett, G., Betz, N. E., Casas, J. M., and Rocha-Singh, I. A. (1992). Gender, ethnicity and social cognitive factors predicting the academic achievement of students in engineering. *Journal of Counseling Psychology* 39: 527-538.

Hartman, H., Everson, H., Tobias, S., and Gourgey, A. (1991). Self concept and metacognition in ethnic minorities. Paper presented at the American Psychological Association Annual Conference, San Francisco.

Haven, E. W. (1971). Factors associated with the selection of advanced academic mathematics courses by girls in high school. Dissertation Abstracts International 32: 1747a, University Microfilms No. 71-26027.

Heller, K. A., and Parsons, J. E. (1981). Sex differences in teachers' evaluative feedback and students' expectations for success in mathematics. *Child Development* 52: 1015-1019.

Higher Education Research Institute (1995). Annual Freshman Survey. Los Angeles: UCLA, Higher Education Research Institute.

Holland, D. C., and Eisenhart, M. A. (1990). *Educated in Romance*. Chicago: University of Chicago Press.

Hyde, J. S., Fennema, E., and Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin* 107: 139-155.

Jacklin, C. (1979). Epilogue. Sex-related differences in cognitive functioning: Developmental Issues. In M. Wittig and A. Petersen (eds.), New York: Academic Press.

Jackson, L. A., Gardner, P. D., and Sullivan, L. A. (1993). Engineering persistence: Past, present, and future factors and gender differences. *Higher Education* 26: 227-246.

Jones, M. G., and Wheatley, J. (1990). Gender differences in teacher-student interactions in science classrooms. *Journal of Research in Science Teaching* 27: 861-874.

Lent, R. W., Lopez, F. G., and Bieschke, K. J. (1991). Mathematics self-efficacy: Sources and relation to science-based career choice. *Journal of Counseling Psychology* 38: 424-430.

Lent, R. W., and Hackett, G. (1987). Career self-efficacy: Empirical status and future directions. *Journal of Vocational Behavior* 30: 347-382.

Lewis, D. J. (1991). Mathematics and women: the undergraduate school and pipeline. *Notices of the American Mathematical Society* 36: 721-722.

Linn, M. C., and Hyde, J. S. (1989). Gender, mathematics, and science. *Educational Researcher* 18: 17-19, 22-27.

Manis, J. D., Thomas, N. G., Sloat, B. F., and Davis, C. S. (1989). *An Analysis of Factors Affecting Choice of Majors in Science, Mathematics, and Engineering at the University of Michigan*. Ann Arbor: University of Michigan, Center for the Education of Women.\newpage\

Marion, S. F. and Coladarci, T. (1993). Gender differences in science course-taking patterns among college undergraduates. Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, Ga. ED363512.

Marsh, H. W., and Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist* 20: 107-123.

Mason, C. L., Kahle, J. B., and Gardner, A. L. (1991). Draw-a-scientist: Future implications. *School Science and Mathematics* 91: 193-198.

Meece, J. L., Parsons, J. E., Kaczala, C. M., Goff, S. R., and Futterman, R. (1982). Sex differences in math achievement: Toward a model of academic choice. *Psychological Bulletin* 91: 324-348.

Mow, S. L. and Nettles, M. T. (1990). Minority student access to, and persistence and performance in, college: A review of the trends and research literature. In J. Smart (ed.) *Higher Education: Handbook of Theory and Research*, Vol. 6. New York: Agathon Press.

Nash, S. C. (1979). Sex role as a mediator of intellectual functioning. In M. A. Wittig and A. C. Petersen (eds.), *Sex-related Differences in Cognitive Functioning: Developmen-*

tal Issues. New York: Academic Press.
National Research Council. (1991). *Women in Science and Engineering: Increasing Their Numbers in the 1990s*. Washington, DC: National Academy Press.
National Research Council. (1994). Survey of Earned Doctorates. Washington, DC: National Academy Press.
Orenstein, P. (1994). *School Girls*. Bantam: New York.
Ottinger, C., and Sikula, R. (1993). Women in higher education: Where do we stand? *Research Briefs* 4: 2. Washington, DC: American Council on Education.
Parsons, J. E., and Adler, T. F. (1983). Expectancies, values and academic behaviors. In J. T. Spence (ed.), *Perspectives on Achievement and Achievement Motivation*. San Francisco: Freeman.
Paulson, K., Hegedorn, L.S., Petrides, L.A., and Wenzel, S.A. (1995). Connecting what we know: the experience of women in engineering education. Paper presented at the Association for the Study of Higher Education, Orlando, Florida, November, 1995.
Post, P., Stewart, M. A., and Smith, P. L. (1991). Self-efficacy, interest, and consideration of math/science and non-math/science occupations among black freshmen. *Journal of Vocational Behavior* 38: 179-86.
Ross, K.C., Volkwein, J.F., and Vogt, W.P. (1995). Women engineering students: Factors associated with their persistence, career commitment, satisfaction, and self-confidence. Paper presented at the Association for the Study of Higher Education, Orlando, Florida, November, 1995.
Rowe, L. M. (1993). Education and occupational sex segregation: the decision to major in engineering. Paper presented at the Annual Meeting of the American Sociological Association, Pittsburgh, PA.
Sadker, D. M., and Sadker, M. (1994). The self-esteem slide. In *Failing at Fairness*. New York: Scribner.
Sarkar, G. (1993). SIAST retention study. Factors affecting retention of first-year students in a Canadian technical institute of applied science and technology. ED356359.
Selvin, P. (1992). Math education: Multiplying the meager numbers. *Science* 258: 1200-1201.
Seymour, E., and Hewitt, N.M. (1994). Talking about leaving: factors contributing to high attrition rates. Final report to the Alfred P. Sloan Foundation. Boulder, CO: Authors.
Seymour, E. (1992). Undergraduate problems with teaching and advising in SME majors--Explaining gender differences in attrition rates. *Journal of College Science Teaching* March/April: 245-254.
Shavelson, R. J., Hubner, J. J., and Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research* 46: 407-441.
Sherman, J., and Fennema, E. (1977). The study of mathematics by high school girls and boys: Related variables. *American Education Research Journal* 14: 159-168.
Sherman, J. (1980). Mathematics, spatial visualization, and related factors: Changes in girls and boys, grades 8-11. *Journal of Educational Psychology* 72: 476-482.
Sikes, J. N. (1971). Differential behavior of male and female teachers with male and female students. *Dissertation Abstracts International*, Volume 217A, University Microfilms, No. 71-1967.
Smith, T. E. (1992). Gender differences in the scientific achievement of adolescents: Effects of age and parental separation. *Social Forces* 71: 469-484.
Tobias, S. (1992). Women in science—Women and science. *Journal of College Science Teaching* Mar/June: 276-78.
Treisman, P. U. (1985). *A Study of the Mathematics Performance of Black Students at the University of California, Berkeley*. Berkeley: University of California.
Ware, N. C. (1988). Sex differences in choice of college science majors. *American Educational Research Journal*: 25 593-614.

Worthley, J. S. (1992). Is science persistence a matter of values? *Psychology of Women Quarterly* 16: 57-68).

Yaeger, P.M. (1995). The "chilly climate for women" revisited: A review of the Literature. Paper presented at the Association for the Study of Higher Education, Orlando, Florida, November, 1995.

Trends in Doctoral Education and Employment[1]

Alan E. Fechter and Catherine D. Gaddy
Commission on Professionals in Science and Technology

INTRODUCTION

An important product of our higher education system is the doctorates it produces. These highly trained human resources create and transmit new knowledge that enhances our productivity, and promulgate cultural values that have and continue to enrich our society. Traditionally, we have had little or no difficulty meeting our need for doctorates in most fields, with the possible exception of science and engineering, where concerns about possible shortages have surfaced periodically. Such concerns surfaced dramatically in the late 1950s, after the Soviet Union launched Sputnik, and gave rise to a concerted effort to recruit a larger share of our workforce to scientific and technological activities.

Concern surfaced again in the late 1980s, when several widely circulated reports forecast shortages of new doctorates (e.g., Bowen and Sosa, 1989; NSF, 1989; NSF, 1990). Subsequent events have invalidated these forecasts and, in more recent times, the concerns have been transformed from those generated by possible shortages to those generated by signs of possible excess supplies (e.g., Massy and Goldman, 1995). There is extensive anecdotal evidence of hundreds of job applicants for tenure-track openings in science and engineering departments (e.g., Ralston, 1996).

These new concerns arise in part from more global changes taking place in our nation. The collapse of the former Soviet Union has fundamentally altered budgetary priorities—especially with respect to the defense budget. International competition has forced American corporations to engage in practices that increase their productivity. Many of these practices have resulted in corporate downsizing and

[1]The authors acknowledge the Science Resources Studies Division of the National Science Foundation (NSF) for support of data compilations specifically for this chapter. In particular, special thanks to Mary Golladay, Carlos Kruytbosch, and Keith Wilkinson of NSF; and Julie Clarke, Susan Mitchell, and Daniel Pasquini of the office of Scientific and Engineering Personnel of the National Research Council. We also sadly report the untimely death of the first author, Alan Fechter, in April 1997 some months after completion of the chapter.

J. C. Smart (ed.), *Higher Education: Handbook of Theory and Research*

cutbacks in hiring. Funding uncertainties at both the federal and the state/local level have made academic institutions more reluctant to undertake long term commitments, such as the hiring of new doctorates into tenure-track appointments.

In fact, the rate of unemployment of new doctorates appears to have increased (e.g., Gaddy, 1995), and there has been a concurrent increase in the percent of new graduates who take temporary positions—e.g. postdoctorates (Simmons and Thurgood, 1995), or other nontenure track appointments. It is not yet clear whether these recent troubling statistics reflect significant structural changes occurring in this particular labor market (as suggested by Tobias, Chubin, and Aylesworth, 1995), or whether this is a more temporary cyclical phenomenon.

This chapter summarizes the pursuit of scholarly activities in the United States. Typically, these activities have primarily involved the production and utilization of doctorates. The purpose is to provide a statistical portrait of doctorates in the United States. It is hoped that, by providing a broad context, such a portrait will help illuminate key issues associated with these recent concerns. Although we try to be comprehensive in our coverage, data limitations force us to concentrate on doctorates in the science/engineering and the humanities fields. In addition, we confine our examination to the past twenty to thirty years.

The next section of this chapter provides an overview of doctorate production in the United States. It provides information on the numbers and demographic composition of this production. It also examines other aspects of the production process, such as the amount of time it takes to complete a doctorate and sources of support while undergoing doctorate training. Finally, it looks at aspects of post-graduation career plans.

Following the section on doctorate production, the chapter turns to an examination of utilization patterns. Rates of unemployment are presented and sectoral patterns of utilization analyzed. In addition, the section provides information on the work activities and salaries of these doctorates.

DOCTORATE RECIPIENTS

For this section of the chapter on doctorate recipients, unless otherwise noted, data are from the Doctorate Records File (DRF), and based on the Survey of Earned Doctorates (SED) (Clarke, 1996). The DRF is a virtually complete database on doctorates recipients from 1920 to the present, that has been maintained to date by the National Research Council with funding from the National Science Foundation, the National Institute of Health, the National Endowment for the Humanities, the U.S. Department of Education, and the U.S. Department of Agriculture. Survey responses to the SED become part of the DRF. For doctorates granted during the 1920-1957 period, information was compiled from commencement bulletins, registrars' records, and other published material (Simmons and Thurgood, 1995).

The SED has been conducted each year since 1958 by the National Research Council. Questionnaires are distributed with the cooperation of graduate deans of U.S. universities, and are completed by graduates as they complete their requirements for doctoral degrees. Response rates from 1964-1994 have been between 92% and 98%. The doctorates are reported by academic year (from July 1 of one year through June 30 of the following year), and include research and applied research doctorates awarded in all fields. Doctoral degrees are covered (e.g., Ph.D.) but professional degrees are not (e.g., M.D.). Additional information on the survey is available in Simmons and Thurgood (1995).

Scope

In this section, overall trends include data for the sciences (physical, life and social), engineering, humanities, education, and professional broad fields. In addition, data are broken out in most sections for (1) science and engineering and (2) humanities, and then within those two broad categories for selected variables in which there is great field-specific variability.

Trends in Numbers

The number of doctoral recipients from U.S. universities has increased over the past 35 years from less than 10,000 to over 40,000. As shown in Figure 1, there was a small dip in the early 1980s, but production increased again thereafter.

In science and engineering fields, as shown in Figure 2, production for selected years has been somewhat uneven for many fields. In physics/astronomy, mathematics, and the social sciences (excluding psychology), production dipped in 1985 but resumed by 1995. For chemistry, environmental sciences (earth, atmospheric, marine) and psychology, growth in production has been relatively gradual over the four decades. Growth has been noticeably greater for the life sciences (biological, health, agricultural) and engineering.

In the humanities, growth in production in all fields decreased in the mid-1980s and for the three specific broad fields shown in Figure 3, the rebound was relatively modest by 1995.

Trends in Composition

Gender. Women increased their share of doctorates over the past 30 years, with a slight tapering off in the most recent decade, as shown in Figure 4. In the humanities, the percentage of women tapered off just below 50% in 1995, while in the social sciences (including psychology), the percentage of women doctoral recipients skimmed past the 50% mark in 1995. The life sciences showed a steady rate of increase. Physical sciences and engineering made steady and comparable gains, with representation of women in engineering growing from less than half a percent in 1965 to almost 12% by 1995.

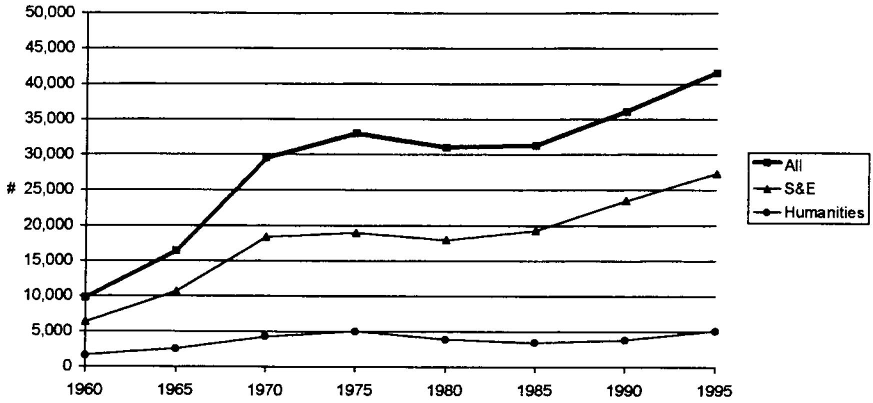


FIGURE 1. Doctorate recipients in the United States: 1960-1995.
(Source: Clarke, 1996)

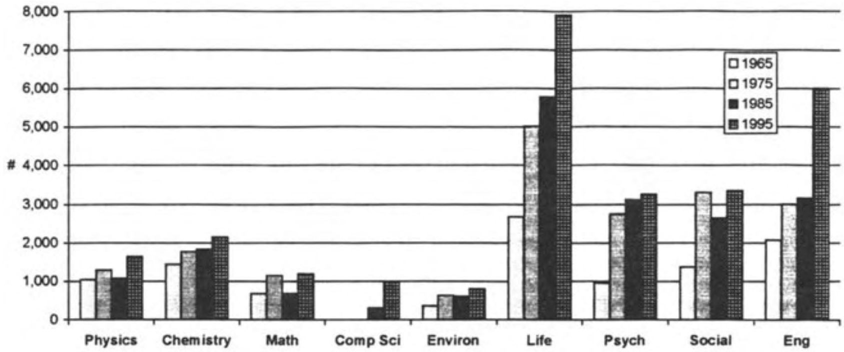


FIGURE 2. Doctorate recipients in science and engineering fields for selected years.
(Source: Clarke, 1996)

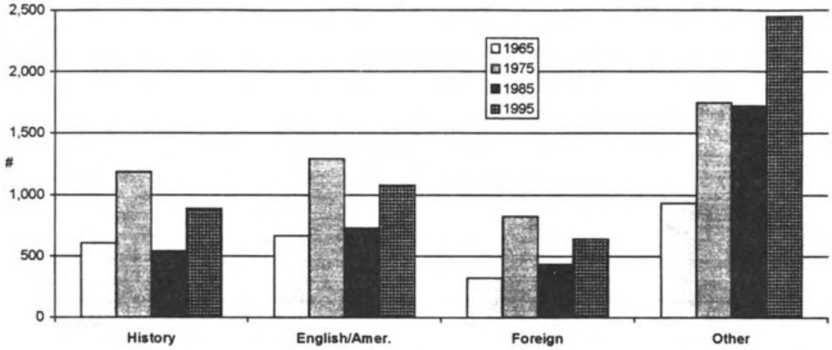


FIGURE 3. Doctorate recipients in humanities fields for selected years.
(Source: Clarke, 1996)

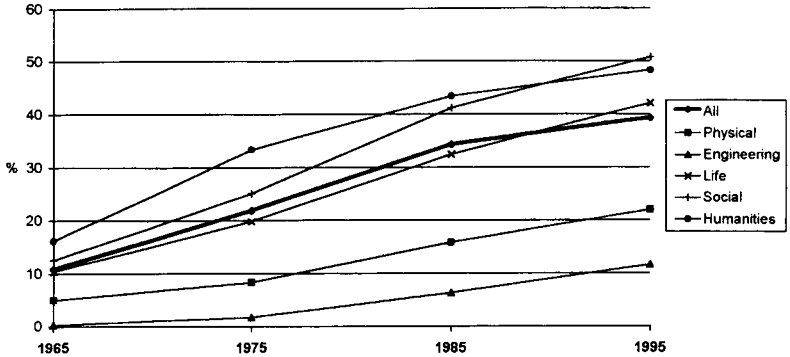


FIGURE 4. Percent female doctorate recipients for selected years. (Source: Clarke, 1996)

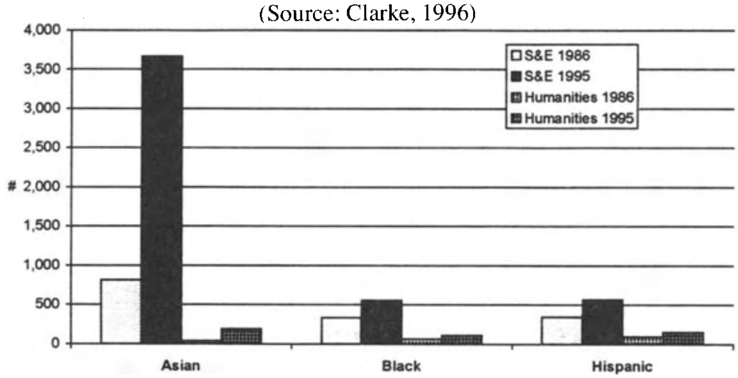


FIGURE 5. Number of doctorate recipients for the three largest minority groups: 1986 and 1995. (Source: NSF, 1996a)

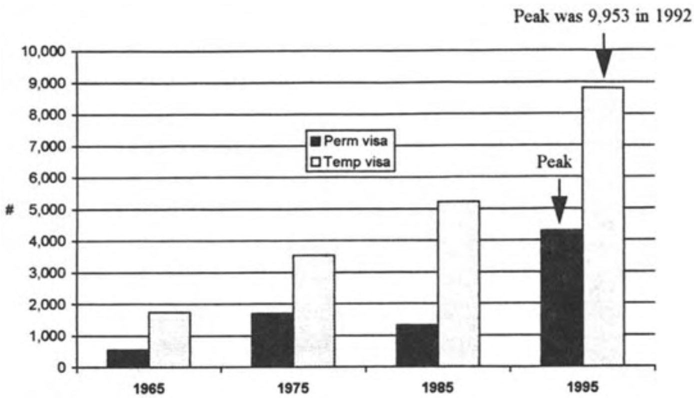


FIGURE 6. Doctorate recipients on permanent and temporary visas for selected years. (Source: Clarke, 1996)

Race/ethnicity. Science and engineering doctorates to Asians increased dramatically between 1986 and 1995 as shown in Figure 5 (NSF, 1996a). *Gains for the two largest underrepresented minority groups, blacks and Hispanics, were much more modest.* Blacks constitute about 12.4% of the U.S. population and Hispanics are about 9.5% (Malcom, George, and Van Horne, 1996). In 1995, the rate of doctoral awards in all fields to Black U.S. citizens was 4.7% and to Hispanic U.S. citizens was about 3.3% (NSF, 1996a).

Citizenship. The number of doctorates awarded to individuals on permanent visas has increased over the past 30 years, and was at an historical high in 1995, as shown in Figure 6. The number of those on temporary visas peaked in 1992 at about 9,950, and has decreased each of the following years, down to about 8,800 in 1995. The distinction between permanent and temporary visas is important, because over 92% of 1995 science and engineering doctorates on permanent visas plan to stay, but only about 57% of those on temporary visas plan to stay (NSF, 1996a). Also, the number of those on temporary visas is much larger than the number on permanent visas.

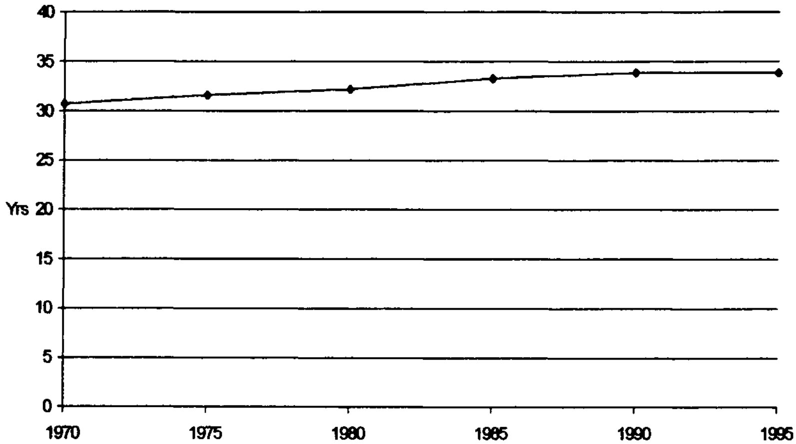


FIGURE 7. Median age at graduation of doctorate recipients for selected years.
(Source: Clarke, 1996)

Age. The median age at graduation has increased since 1970 as shown in Figure 7. In 1970, the median age for doctoral recipients was 30.7, and the age had increased to almost 34 by 1995. The youngest doctorates were in the physical sciences, at 31 years old, but this too was an increase over the median age of 28 in 1970. Other median ages in 1995 were: 31.7 for engineering, 32.4 for life sciences, 34.1 for social sciences, and 35.4 for humanities. All the science graduates were at or near 30 in 1970. The time taken to complete the degree is responsible to some extent for this "aging" of the graduates, as described in the section that follows.

Trends in Support and Time to Degree

Sources of support. Trends in financial support for doctoral graduate education indicate that the percentage of those using personal resources has decreased recently. In 1977, it was 36.1%; it peaked at 42.8% in 1984, and decreased to 35.8% in 1995.

Graduating doctorates reporting university support increased during the same time period from 41.9% in 1977 to 52.5% in 1995. It is important to note that research assistantships funded by the Federal government, primarily via research grants to faculty, are counted as university support (Simmons and Thurgood, 1995).

Direct Federal support to these doctoral students (in forms such as fellowships from a source such as NIH or NSF directly to the individual student) decreased from its highest point (according to available data), at 16.1% in 1977 down to 5.6% in 1995. Thus, Federal policy appears to have changed, with university support increasing in importance and fellowships decreasing in importance. "Other" sources ranged between 6-7% over the period.

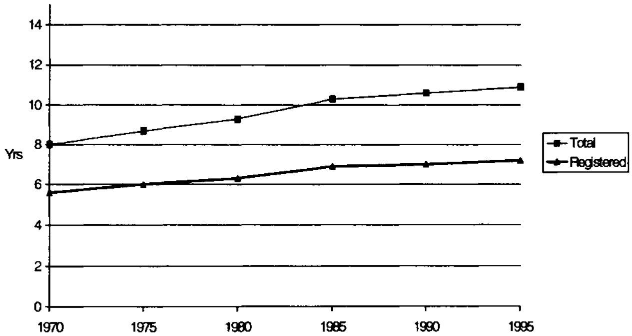


FIGURE 8. Median time to degree for doctorate recipients for selected years.
(Source: Clarke, 1996)

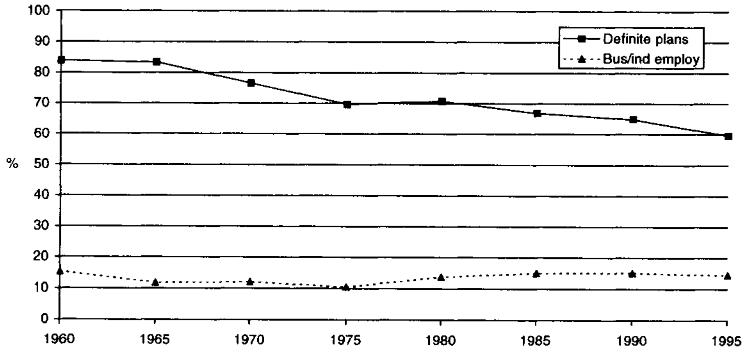


FIGURE 9. Percentages of doctorate recipients with definite plans and with definite plans for employment in business and industry for selected years.
(Source: Clarke, 1996)

Time to degree. Both total (i.e., time from baccalaureate) and registered time to obtain the doctoral degree increased between 1970 and 1995, as shown in Figure 8. In 1995, it took the average doctorate 1.6 additional registered years to graduate than it had in 1970. In 1995, while engineering (6.4 years) and physical sciences (6.9 years) had the lowest registered times to degree, these times were still a year to a year and a half greater than in 1970. Life science doctorates took 7.0 registered years in 1995 compared with 5.3 years in 1970. Registered time for completion of social sciences doctorates increased by 2.0 years (to 7.5), and humanities by 2.4 years (to 8.4).

Trends in Postgraduation Transitions

Postgraduation employment plans. Doctoral recipients of late have less certain immediate futures than the 1960s cohorts. As shown in Figure 9, the percentage of doctoral graduates with definite postgraduation plans has decreased steadily from 84% in 1960 to just below 60% in 1995. Those who planned to go on to temporary postdoctoral study positions increased from 8.6% in 1960 to 29.8% by 1995. At the high end of the range in 1995, of those doctoral graduates remaining in the U.S., 76.5% of those in the biological sciences who had definite plans were headed for postdoctoral positions (NSF, 1996a). The average rate for science and engineering overall, fields in which postdocs are most prevalent, was 41% in 1995 (NSF, 1996a).

During the same time period, the percentage of new graduates with definite plans to work in educational institutions immediately after graduation decreased from 57.2% in 1960 to 43.4% in 1995. The percentage with definite plans to work in business/industry immediately after graduation also decreased, from 15.4% in 1960 to 14.5% in 1995. However, the trend in plans of recent doctorates to go to work in business/industry has not been constant over time. As shown in Figure 9, the percentage of new doctorates with definite plans to go to work in business/industry immediately after graduation has varied between 10% and 15%. According to the definite plans of graduates in 1995—in the physical sciences, 21% were bound for industry; in engineering, 44.7%; in life sciences, 7.3%; in social sciences, 11.9%; and in humanities, 4.7%

First employment experiences of recent graduates. Four professional societies/ associations, the American Chemical Society (ACS), the American Institute of Physics (AIP), the American Mathematical Society (AMS), and the American Psychological Association (e.g., Wicherski and Kohout, 1995), have collected employment data from recent doctoral graduates up to about a year after graduation. Three of the societies (ACS, AIP, AMS) reported increases in the percentage of doctoral graduates who were still unemployed at the time of their surveys (ACS, 1995; AIP, 1996a; Czujko and Dodge, 1996; Fulton, 1996). As shown in Figure 10, the magnitude of the trends varies across the fields, but the increases are evident for all three. The most recent data available from AIP suggest that the trend for physics may be turning around (AIP, 1996b).

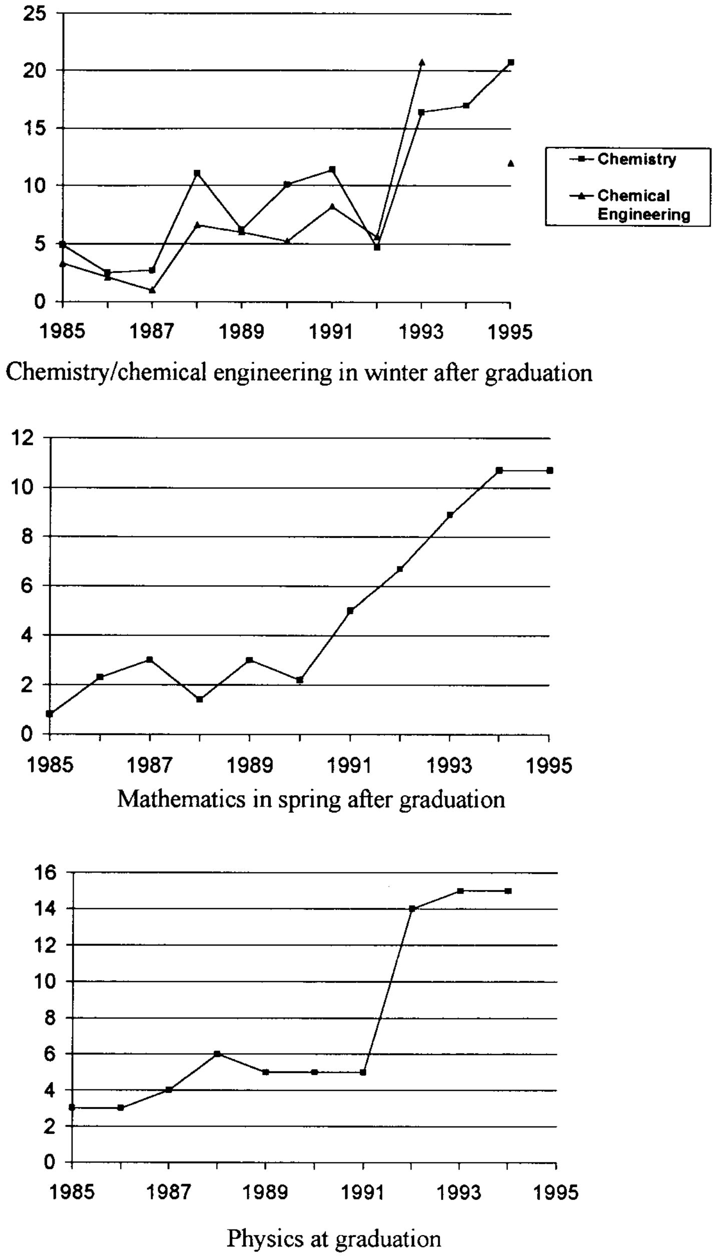


FIGURE 10. Percentages of new doctorates unemployed and looking.
(Sources: ACS, 1995; Czujko and Dodge, 1996; Fulton, 1996)
Note: Data not available for chemical engineering for 1994.

The Commission on Professionals in Science and Technology (a participating organization of the American Association for the Advancement of Science), in collaboration with these four and other professional societies, has been funded by the Alfred P. Sloan Foundation and the National Science Foundation to increase the availability and consistency of such data for science and engineering fields.

EXPERIENCED DOCTORATES

Recent data describing the characteristics of the doctorate population in the United States are available from the biennial Survey of Doctorate Recipients (SDR). These characteristics are estimated from a stratified random sample of doctorate holders.

The SDR was first conducted in 1973 for doctorates in fields of computer and mathematical sciences, life and related sciences, physical and related sciences, social and related science, and engineering. It was expanded in 1977 to include humanities doctorates in the broad fields of history, art history, music, philosophy, English/American language/literature, classics, and modern language/literature.

The primary sampling frame for the survey is the Doctorate Records File (DRF), which was described earlier. The sample is updated biennially by adding the two most recent graduating cohorts and removing the two oldest cohorts. Members of the sample are then contacted by mail and asked to provide information on their employment and sociodemographic characteristics.

In 1991, the methodology of data collection was changed, and telephone follow up was introduced to increase the response rate. In 1993, the survey instrument went through a major redesign. As a result, data from previous years are not directly comparable. Thus, trend comparisons between these years and periods prior to 1991 should be treated with caution because of possible biases introduced by these methodological changes. These biases will be more serious for recent comparisons than they will be for longer range comparisons. For this reason, data in the sections that follow will be presented in time series to illustrate historical trends up through 1989 (NSF, 1991; Brown and Pasquini, 1991), with snapshots of data for 1991 (NSF, 1994; Brown and Mitchell, 1994) and 1993 (NSF, 1996b; Ingram, Brown, and Mitchell, 1995) presented separately. The analysis focuses on the employed population since relatively few doctorates are unemployed or out of the labor force, and there is no apparent trend in unemployment rates or labor force participation rates.

Trends in Numbers

The number of employed doctoral scientists and engineers approximately doubled from 1973 through 1989, rising from over 220,000 to almost 449,000, as shown in Figure 11. The rapid growth in employment of doctorates suggests

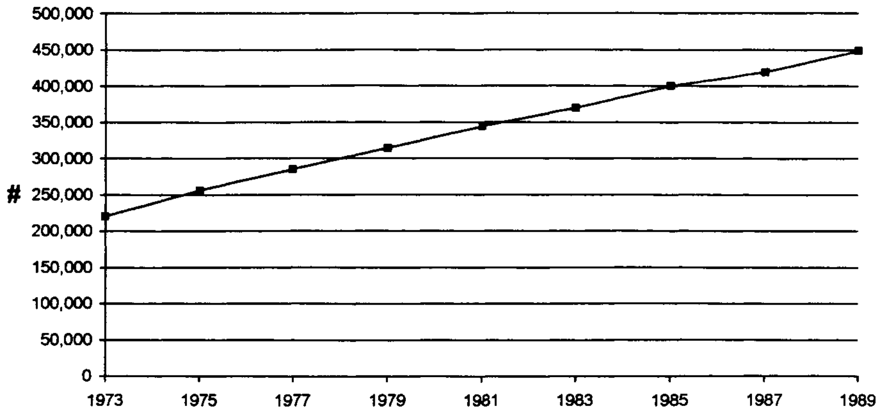


FIGURE 11. Employed doctoral scientists and engineers: 1973-1989.
(Source: NSF, 1991)

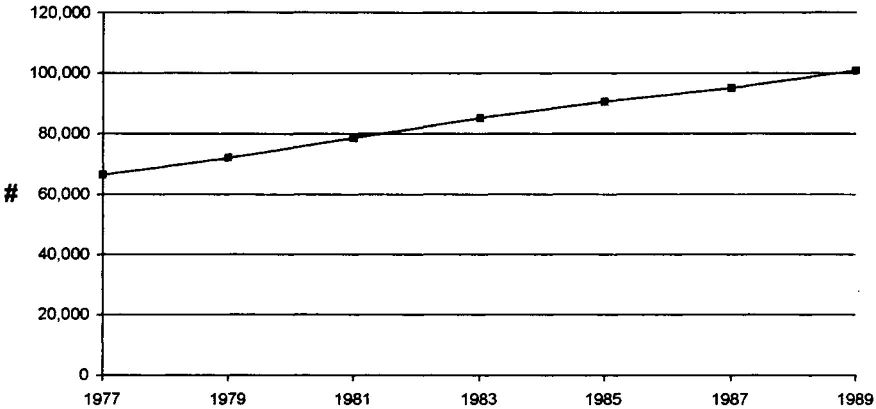


FIGURE 12. Population of humanities PhDs: 1977-1989.
(Source: Brown and Pasquini, 1991)

demand for these skills was increasing over this period. By 1991, the number of employed doctoral scientists and engineers was almost 486,000; and by 1993, it was over 513,000.

The population of humanities doctorates is profiled in Figure 12. A more modest rate of increase is shown from a smaller base. In 1991, the population was estimated to be 100,300; and by 1993, it was 106,200.

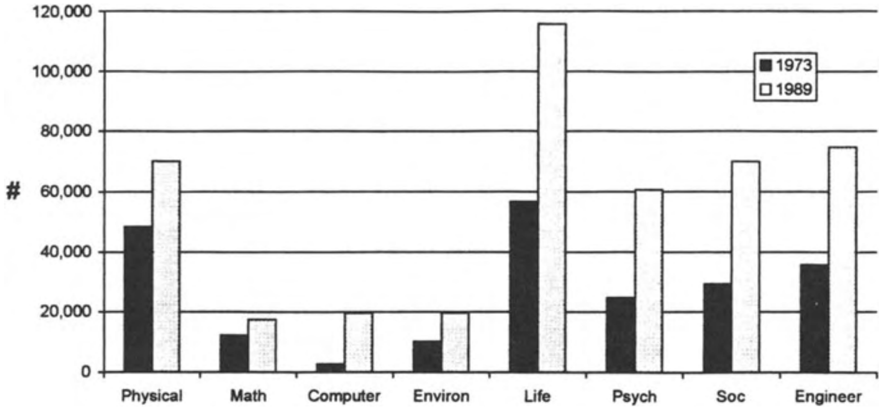


FIGURE 13. Employed doctoral scientists and engineers by broad field: 1973/1989.
(Source: NSF, 1991)

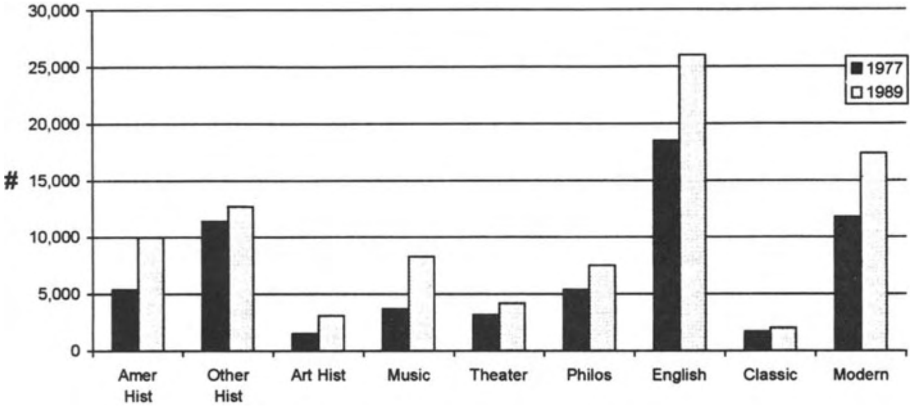


FIGURE 14. Population of humanities PhDs by field: 1977/1989.
(Source: Brown and Pasquini, 1991)

By field within science and engineering, as shown in Figure 13, all the broad fields grew in numbers of employed doctorates. The life sciences remained the largest field in 1989, as it was in 1973. The physical sciences were in second place in terms of numbers of employed doctorates in 1973, but by 1989, engineering had grown somewhat larger with the social sciences not far behind. As would be expected given technology trends, the number of doctoral scientists in the field of computer sciences increased significantly over the time period.

Within the humanities, all the broad fields increased in size as well, as shown in Figure 14. English/American language and literature remained the largest broad field in terms of numbers of doctorates, with modern language and literature remaining in second place. Other history (other than American) ranked third in terms of size. The number of humanists in American history almost doubled in size. In art history, the number doubled from a relatively small base. In music, the number more than doubled. Speech/theater and philosophy showed gains as well. Classical language and literature grew modestly from its relatively small base.

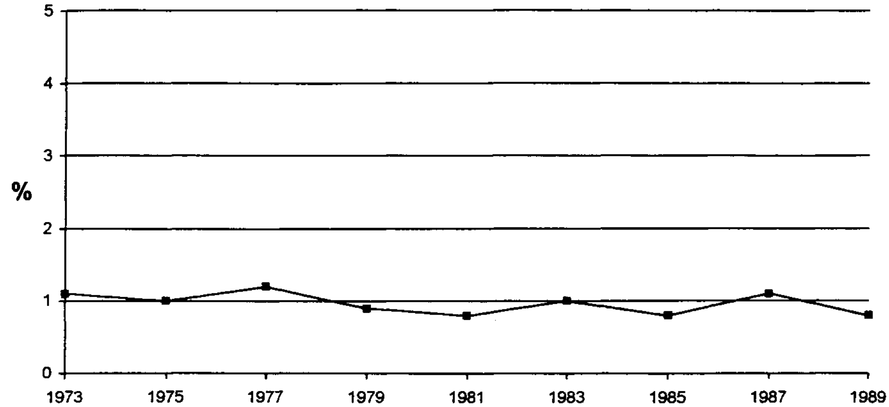


FIGURE 15. Unemployment rate for doctoral scientists and engineers: 1973-1989.
(Source: NSF, 1991)

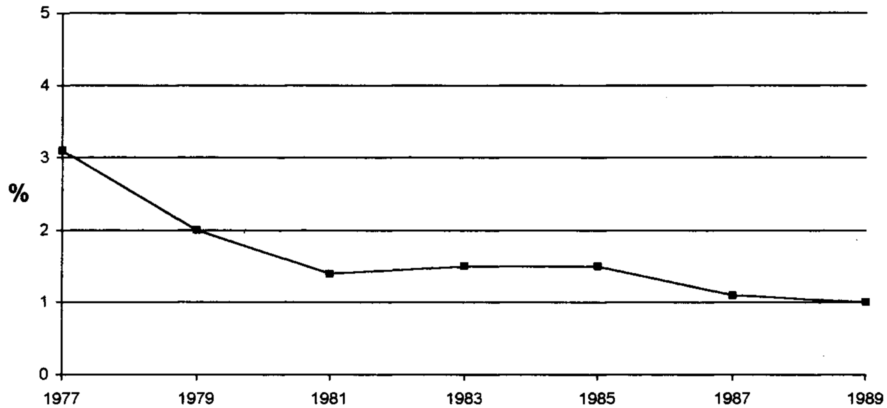


FIGURE 16. Unemployment rate for humanities PhDs: 1977-1989.
(Source: Brown and Pasquini, 1991)

Trends in Unemployment

The unemployment rate is one measure of labor market conditions. If the rate is relatively high, there may be weak demand and an unbalanced market; if it is low, there may be strong demand and a balanced market. Unemployment rates for doctoral scientists and engineers ranged between 0.8% and 1.2%, with no particular pattern of increase or decrease over the years, as shown in Figure 15. The unemployed are defined for the purposes of these analyses as those who reported that they were not employed and were *seeking* employment. In 1989, the overall rate was 0.8%. Women had more than twice the rate of men at 1.7% compared with 0.6%. Blacks had the highest rate (3.7%) of all the racial groups. Doctorates under the age of 30 had a higher unemployment rate than other age cohorts, although this may be attributed to their lesser experience. Within fields in 1989, sociologists/ anthropologists had the highest unemployment rate at 2.9%, while materials science engineering had the lowest at 0.1%.

In 1991, the unemployment rate for doctoral scientists and engineers was 1.4% (NSF, 1994), and it was 1.6% in 1993 (NSF, 1996b). (Although there are methodological differences in the calculations, for rough comparison purposes, the annual average unemployment rate for the entire U.S. workforce for 1994 was 6.1%, and for 1995 it was 5.6% (U. S. Department of Labor, 1996).) Earlier differences in unemployment rates by sex and race/ethnicity were not apparent in 1993. The highest rate of unemployment was 3.0% for those with doctoral degrees in geology and oceanography. The lowest was a tie between computer and information sciences and mechanical engineering. The percent of those employed involuntarily out of their field was 4.3% for all doctoral scientists and engineers.

In the humanities, there was a downward trend in unemployment rates from 1977 through 1981, as shown in Figure 16. After that, unemployment rates for humanists and for scientists and engineers were roughly comparable. The rate decreased from 3.1% in 1977 to 1.0% by 1989. In both 1991 and 1993, the rate was 1.7%.

In 1993, unemployment within the humanities was highest for those in art history, both for all humanists in this field and for recent art history graduates. Lowest unemployment was reported by those in modern language/literature, other humanities and music, again both for all humanists in these fields and for recent graduates. In all fields, recent graduates reported higher unemployment than for all humanists as a group, ranging from about 1% higher unemployment for recent graduates in music to over 3% higher for recent graduates in philosophy.

Trends in Employment

By Employment Sector. The percent of doctoral scientists and engineers employed in business/industry rose dramatically between 1973 and 1989. At the same time, the percent employed in academe decreased, as shown in Figure 17.

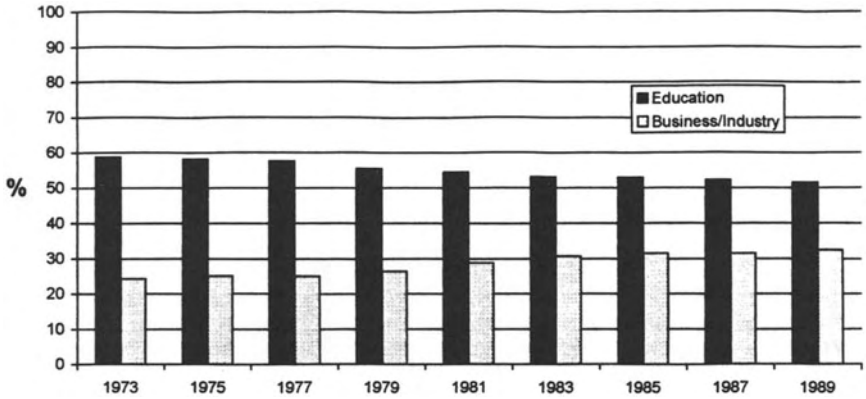


FIGURE 17. Major employment sectors of doctoral scientists and engineers: 1973-1989. (Source: NSF, 1991)

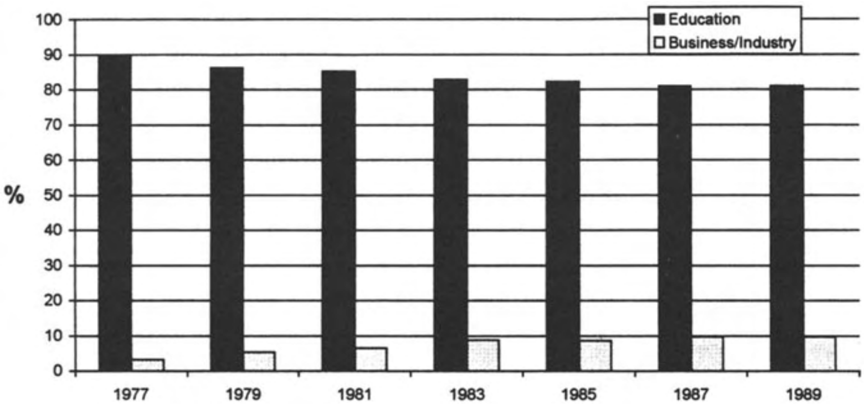


FIGURE 18. Major employment sectors of humanities PhDs: 1977-1989. (Source: Brown and Pasquini, 1991)

In 1991, employment in the education sector dropped below 50% to 47%, while employment in business/industry increased to 36%. One must keep in mind that the use of telephone follow-ups may have resulted in a different sample than in previous years. Still, a trend toward greater employment in business/industry seems to be a real phenomenon, but probably is occurring at a pace that appears more dramatic than is actually the case because of changes in survey methodology. In 1993, the percent employed in education was 48% (including less than 3% not in colleges/universities), and was 31% in private/for-profit organizations, 6% self employed and 5% private/not-for-profit organizations.

This trend is important because traditionally the focus of graduate education and training has been to train doctoral scientists and engineers for positions in research universities. While this will undoubtedly remain a major goal for some doctoral programs, faculty may need to consider implications for education/training as an increasing percentage of doctorates pursue opportunities outside academe (COSEPUP, 1995; Edwards, 1990; Klatzky, 1990; Tobias, Chubin, and Aylesworth, 1995).

Data on the employment pursuits of doctorates in non-academic settings that could inform academic decision-making are more difficult to develop because of differences in corporate settings. Moreover, some of these pursuits are taking place outside of industrial research laboratories. There is not a great deal of information on these less traditional career pursuits, as primarily anecdotal evidence of careers in finance and other business-related endeavors has accumulated. Skill sets, knowledge needed, and forms of accomplishment vary widely across sectors and across career life spans.

However, information about these new types of opportunities can broaden and enrich both the educational experiences and careers of doctorates in science and engineering. And, spreading science and engineering further beyond the walls of academe has great potential for expanding the base of support for the generation of knowledge, within academe and outside, to support diverse applications.

A similar trend is observed among humanists, with an almost 10% decrease in the percentage of those employed in education, as shown in Figure 18. A concurrent increase occurred in the percentage employed in business/industry, but still represented only about 10% of doctoral humanists by 1989. Thus, education was still the dominant sector of employment.

By 1991, the fraction employed in education decreased slightly to 78%, while the fraction in business/industry increased slightly to 12%. In 1993, 79% were in education, with 7% in private/for-profit, 5% self employed, and 6% in private/not-for-profit.

If the job market in academe is adversely impacted in comparison with other sectors, humanists may feel more of the effects than their colleagues in science and engineering, because of their greater dependence on this sector. At the moment, however, it appears that the job market in all sectors is facing challenges because of the economy and other factors discussed earlier, so diversity in sectors of employment may not be as viable a solution to finding good employment matches for doctorates as it might have been in earlier times when sectors tended to have downturns at different times.

By Type of Work Activity. As shown in Figure 19, scientists and engineers tend to have diverse primary work activities including teaching, research and development and management/administration. Teaching predominates as the primary work activity of humanists consistently over time.

For faculty in doctoral programs in the humanities, the challenges in preparing a new generation of doctorates may come less from diversification of employment sector (see Figure 20), as is the case for educators of scientists and engineers, and more from the rapid developments in technology affecting education. As colleges and universities develop increased capabilities to teach via a broad array of computer technologies, faculty in doctoral programs will need to prepare future teachers to use these new approaches as best they can, given the dizzying speed with which they are evolving.

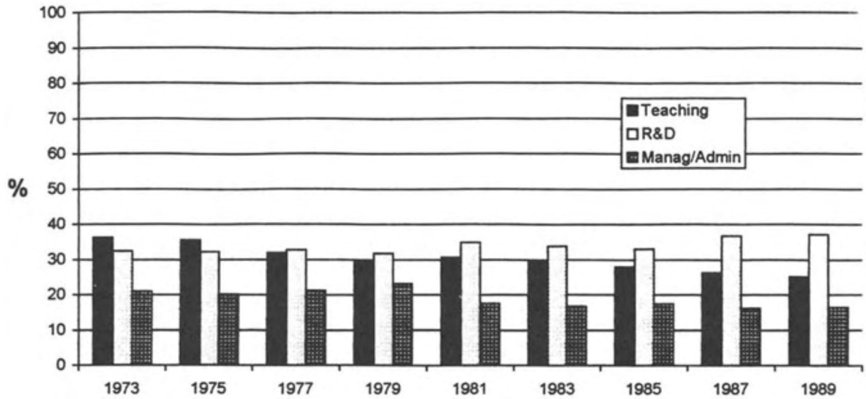


FIGURE 19. Primary work activities of doctoral scientists and engineers: 1973-1989.
(Source: NSF, 1991)

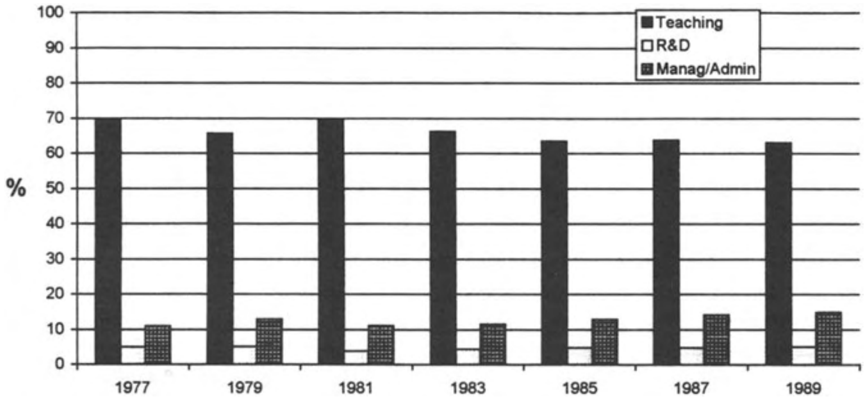


FIGURE 20. Primary work activities of humanities PhDs: 1977-1989.
(Source: Brown and Pasquini, 1991)

Trends in Salaries. Salaries for doctorates in science/engineering and the humanities are illustrated in Figures 21 and 22, respectively. For scientists and engineers, salaries have increased at a greater rate than for humanists over the period from 1977 to 1989. One reason for this difference is the increasingly higher percentages of scientists and engineers employed in business/industry, which typically offers higher salaries than in education (e.g., the median salary for scientists and engineers in colleges/universities in 1993 was $51,100, while in the private/for-profit sector it was $71,900) (NSF, 1996b).

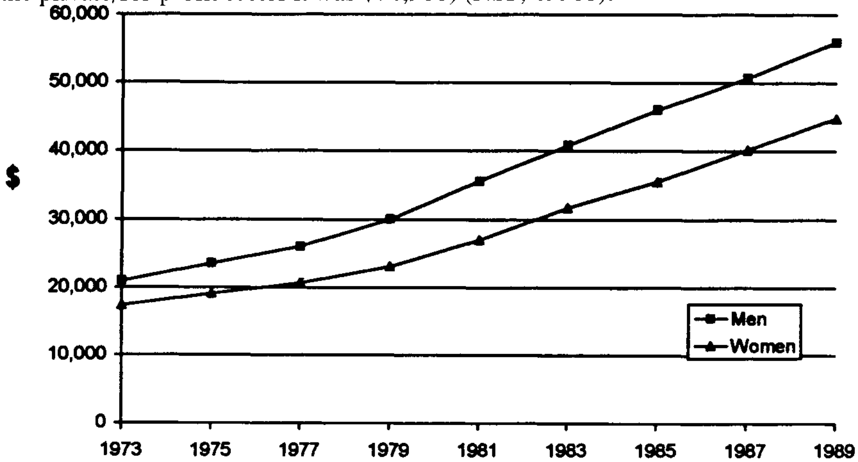


FIGURE 21: Median salaries for doctoral scientists and engineers by gender: 1973-1989. (Source: NSF, 1991)

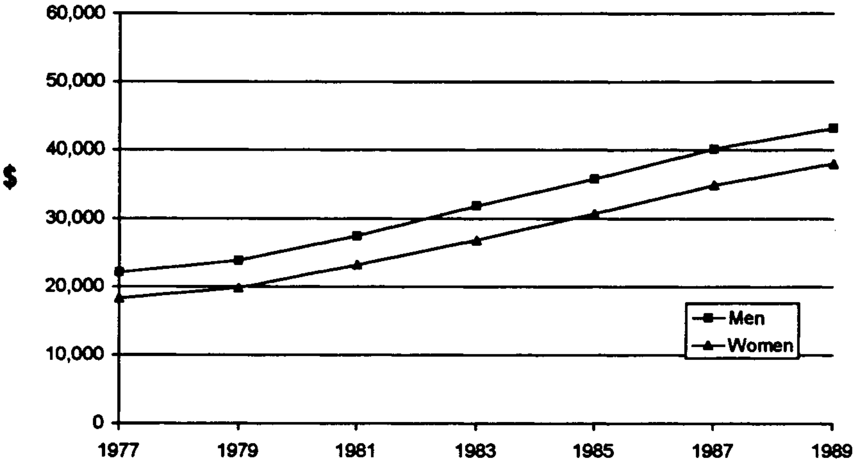


FIGURE 22. Median salaries for humanities PhDs by gender: 1977-1989. (Sources: Belisle and Brown, 1989; Brown and Pasquini, 1991; Maxfield, 1980, 1982; Maxfield, Ahern, and Spisak, 1978; Maxfield and Brown, 1985, 1986)

With respect to gender differences in salaries, the difference between median salaries for men and women widened in science and engineering over the same time period. In 1993, men's median salary was $61,500 compared to $48,400 for women (NSF, 1996b). In part, this was due to an influx of less experienced women into fields historically (and still) dominated by men, such as chemistry and physics.

However, analyses that control for factors such as employment sector or years experience typically still find a small gap that cannot explained by these types of variance alone. For those who worked full time and were not self-employed, the average salaries were $50,200 for women compared with $63,600 for men (NSF, 1996c). The salary gap of $13,300 corresponds to women making 79% of men's salaries. When variables including year of doctorate, degree field, employer and type of work were held constant, these variables accounted for an estimated $11,900 of the difference (NSF, 1996c).

In the humanities, a relatively consistent discrepancy between men's and women's salaries is found. Similar gaps are found relatively consistently across all the fields as well (Ingram, Brown and Mitchell, 1995). In terms of career age, in 1993, younger women tended to earn about the same as younger men, but older women earned less than older men (Ingram, Brown and Mitchell, 1995). The issue of salary equity for doctorates in all fields is one that remains to be addressed fully.

Trends in Productivity/Achievement. No data were collected as part of the SDR up through 1989 on presentations/publications, patents, performances or other achievements. However some crude measures of productivity or achievement are available, primarily in the forms of numbers of and citations for publications, and numbers of patents. A broader set of outcome measures to gauge achievements of doctorates may be needed (e.g., Richards and Gottfredson, 1984), particularly for those doctorates whose education or research has been supported by Federal funds.

Based on the Institute for Scientific Information's Science Citation Index (NSB, 1996), the number of U.S. scientific publications rose from about 132,000 in 1981 to 141,000 in 1993 (an increase of about 7%, during a time in which the numbers of doctoral scientists and engineers increased about 41%—from 364,000 to 513,000). At the same time, the U.S. contribution of publications to the worldwide literature decreased from 36% in 1981 to 34% in 1993, reflecting the relatively more rapid growth in publications output of other nations such as China, Hong Kong, Japan, Singapore, South Korea, and Taiwan. The increasing globalization of science and engineering research will have varied impacts, including increased need to understand the conduct of science in different cultures and increased opportunities for multinational collaboration.

Counting whether one published or not, or how many articles one published, is obviously a relatively crude measure of achievement, but it is one approach. Citations provide another potential measure that can provide information on the usefulness of publications as judged by other researchers.

As part of a study of former recipients of prestigious postdoctoral positions primarily from the National Research Council and National Science Foundation, Sonnert (1995-1996) found that men published an average of 2.8 scientific publications per year and women published 2.3. Sonnert and Holton (1996) also conducted a small sub-study of 25 former NSF fellows in biology and examined citations in the scientific literature to these biologists' articles. They found that articles written by the women in this small sample were cited on average 24 times compared with an average of 14 citations per article for the men. While their sample was small, it reminds us of the need to consider relatively sophisticated measures of achievement lest we oversimplify them.

The number of academic institutions granted patents increased from about 80 in 1980 to about 160 in 1994. At the same time, the number of patents granted to US academic institutions increased from about 400 in 1980 to about 1,700 in 1994 (NSB, 1996). The actual number of scientists and engineers who contributed to these efforts is not provided. Further, patents in business/industry to which doctoral scientists and engineers made contributions historically are more difficult to identify. This type of achievement data, in addition to presentations/publications, was collected beginning with the 1995 SDR for scientists and engineers (data are not yet available).

In 1991, 69% of doctoral humanists had published one or more times in the preceding *two* years (Brown and Mitchell, 1994). In 1993, over half (57%) of humanists had a publication in the *most recent year* (Ingram, Brown and Mitchell, 1995). Historians and art historians were more likely to have published than humanists in other fields. Data on exhibitions or performances in the fine or applied arts were collected but not reported.

Trends in Funding Support. The percentage of doctoral scientists and engineers who received Federal support for their work dipped in 1985 after a gradual decline, but recovered in the late 1980s, as shown in Figure 23. By 1991, 41% had received some form of Federal support in the *preceding year* (NSF, 1991). In 1993, 26% reported support during the *week of April 15* (NSF, 1996b). (This provides a good example of the differences that may result based on changes in the content of the survey.)

Substantially fewer humanists received Federal support. In 1991, about 15% of humanities doctorates reported being engaged in work supported or sponsored by the Federal government in the *preceding year* (Brown and Mitchell, 1994). In 1993, during the *week of April 15*, 5% received some type of support, including 8% of those in art history (the second smallest field in terms of numbers), followed by history with about 6% (Ingram, Brown and Mitchell, 1995).

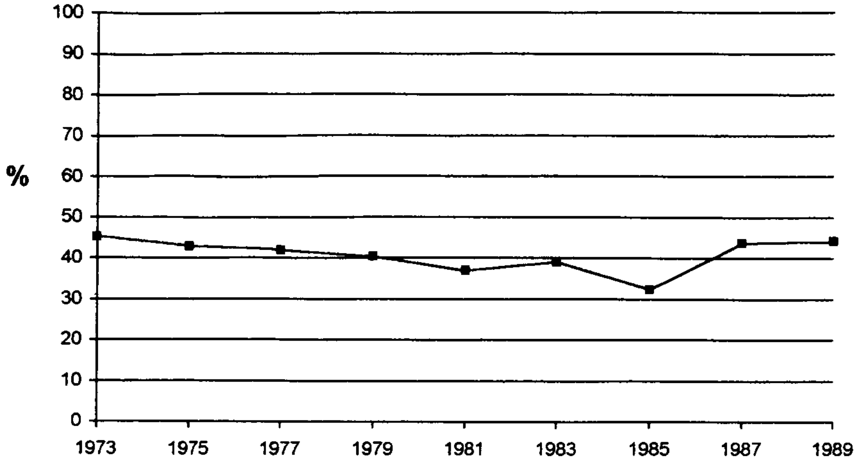


FIGURE 23. Percentages of doctoral scientists and engineers with Federal support: 1973-1989. (Source: NSF, 1991)

SUMMARY AND CONCLUSIONS

U.S. universities are producing record numbers of doctorates at the present time. Although there was a slight decrease in production in the early 1980s, the numbers increased again, and set a record in 1995. Between 1965 and 1995, all major fields within science and engineering increased production, with the most substantial increases occurring in the life sciences and engineering, each of which tripled production. Growth was evident in major fields within the humanities, although fields such as history, English/American language/literature, and foreign language/literature had not returned to the 1970s levels of production by 1995. On the positive side, it appears that the relatively lower production in the humanities during that time period resulted in a concomitant decrease in unemployment.

Representation of women was at parity with men in the social sciences and humanities by 1995, while the life sciences were getting closer to parity. Gains were made in doctoral awards to women in the physical sciences and engineering, but by 1995, women were still less than a quarter of the awardees in the physical sciences and just over 10% in engineering.

In terms of minority representation, Asians quadrupled the number of doctorates they received between 1986 and 1995. However, gains for the two largest, underrepresented minority groups were not as impressive. While Blacks constitute about 12% of the population, they are only about 5% of doctorates in science and engineering, and are represented even less in the humanities. Likewise, Hispanics comprise nearly 10% of the U.S. population, but receive only about 3% of doctorates. In an era in which affirmative action is being challenged (Malcom, George, and Van Horne, 1996), higher education may need to address diversity in doctoral education in new and creative ways.

The number of doctorates awarded to individuals on temporary visas peaked in 1992 and has decreased each of the following years. Those on permanent visas continue to increase, but this is a much smaller number of individuals, at just over 4,000 in 1995. As many other countries continue to increase doctoral production internally (NSB, 1996), the U.S. may see the numbers of those on temporary visas continue to decrease.

New doctorates are older at graduation than earlier cohorts. The median age increased by over three years between 1970 and 1995, due in part to increases in time to degree of a little over a year and a half during the same time period. Time-to-degree has numerous facets and implications (Bowen and Rudenstine, 1992). Among them, the attractiveness of the fields of study for prospective students may become an issue at some future time, if the investment is perceived as too great for the benefit.

In comparison with 1977, doctorates received proportionately somewhat more financial support from universities in 1995, than from either personal resources or direct Federal support. However, it is likely that much of the university support to students is received from the Federal government in the form of research grants to faculty, who in turn employ research assistants. Fewer than 6% of recent graduates had received direct Federal support to finance their education. Issues associated with different forms of support and their implications for subsequent career development are topics of current debate within the science community.

Doctorates who have graduated recently have faced a less certain immediate future than earlier cohorts. While 84% of the class of 1960 had definite postgraduation plans, by 1995, less than 60% knew what was coming next. At the same time, record numbers of graduates, especially in the biological sciences were headed for temporary postdoctoral positions. While current employment statistics indicate that doctorates eventually find employment, the uncertainty and time required in the school-to-work transition are issues for faculty, employers, and others to bear in mind.

In a related vein, in follow-up surveys conducted by professional societies after graduation, increasing percentages of doctorates in chemistry, mathematics and physics in recent years have reported that they have not yet found employment within about six months to a year after graduation. Comparable data are not available for other fields, but an NSF- and Sloan-funded CPST/professional societies initiative is underway to increase the availability and comparability of data for additional fields.

As would be expected given the production numbers, the numbers of all doctoral scientists, engineers, and humanists in the U.S. continues to increase. Between 1973 and 1993, the number of employed doctoral scientists and engineers more than doubled from about 220,000 to over 513,000. The greatest growth occurred in the life sciences and engineering, reflecting the relatively large increases in production. Doctoral humanists increased steadily in number to exceed 106,000 by 1993. English continued to lead as the largest of the major fields, while modern languages increased substantially, and American history and music each doubled in the sizes of their populations.

Unemployment rates for doctoral scientists/engineers and for humanists were at about the same relatively low level of 1% in 1989. For scientists and engineers, this was a continuation of a flat "trend," while for humanists this was a marked improvement over the more than 3% unemployment of the mid-1970s. The rates for both scientists/engineers and humanists had edged up to about 1.6-1.7% in 1993, so this statistic bears monitoring.

Increasing percentages of doctoral scientists and engineers worked in business/industry with decreasing percentages working in educational institutions from 1973 to 1989, and this trend seems to be continuing into the 1990s. A gradual shift in primary work activity away from teaching and toward research and development accompanied this sectoral shift. Almost 10% fewer humanities doctorates worked in education in 1989 compared with 1977, but education was still the sector of employment for over 80% of the humanities doctoral workforce, with teaching the predominant primary work activity.

Salaries for scientists/engineers were greater on average than for humanists, which is accounted for in part by the sectoral differences in employment. Salary gaps between men and women were found, and at least some of the difference could be accounted for by the younger age of the women in the populations and their choices of field. However, unexplained gender gaps still remained.

Data on productivity measures for science and engineering doctorates (e.g., publications and patents), collected recently for NSF as part of the SDR, will provide a start at further describing accomplishments these doctorates. In the meantime, data on international scientific publication indicate that publications of scientists in other countries are on the increase, demonstrating further the globalization of the research enterprise. On the home front, patents granted to academic institutions more than quadrupled between 1980 and 1994, and certainly the contributions of doctoral scientists and engineers to patents awarded in business/industry are sizable. Almost 60% of humanists had published in 1993, with historians and art historians leading the way. Accomplishments in the forms of exhibitions and performances add to their contributions. One fruitful avenue for further research is the pursuit of additional outcomes such as these to allow doctorates to gauge their personal success, and to allow society to better understand the contributions of this highly educated segment within our society.

References

American Chemical Society. (1995). *Starting Salaries of Chemists 1995*. Washington, D.C. (and previous annual editions back to 1985)

American Institute of Physics. (1996a, May). 1995 initial employment follow-up of 1994 physics degree recipients. *AIP Report* No. R-282.18. College Park, MD: Author.

American Institute of Physics. (1996b, November). 1996 initial employment report: Follow-up of the 1995 physics degree recipients. *AIP Report* No. R-282.19a. College Park, MD: Author.

Belisle, M., and Brown, P. (1989). *Humanities Doctorates in the United States: 1987 Profile*. Washington, DC: National Academy Press.

Bowen, W. G., and Rudenstine, N. L. (1992). *In Pursuit of the Ph.d.* Princeton, NJ: Princeton University Press.

Bowen, W. G., and Sosa, J. A. (1989). *Prospects for Faculty in the Arts and Sciences: A Study of Factors Affecting Demand and Supply, 1987-2012*. Princeton NJ: Princeton University Press.

Brown, P., and Mitchell, S. (1994). *Humanities Doctorates in the United States: 1991 Profile*. Washington, DC: National Academy Press.

Brown, P., and Pasquini, D. (1991). *Humanities doctorates in the United States: 1989 profile*. Washington, DC: National Academy Press. (Contains selected time series data from 1977-1989)

Clarke, J. E. (1996). Special analyses based on the Doctoral Records File, Survey of Earned Doctorates. Washington, DC: National Research Council.

Committee on Science, Engineering, and Public Policy. (1995). *Reshaping the Graduate Education of Scientists and Engineers*. Washington, DC: National Academy Press.

Czujko, R. and Dodge, B. (1996). Unpublished data table of historical unemployment rates for recent doctorates in physics. College Park, MD: American Institute of Physics.

Edwards, D. C. (1990). Should departments be held responsible for the marketability of their graduates? A con position. In L. Bickman and H. Ellis (eds.), *Preparing Psychologists for the 21st Century: Proceedings of the National Conference on Graduate Education In Psychology*. Hillsdale, NJ: Lawrence Erlbaum.

Fulton, J. D. (1996, August). 1995 Annual AMS-IMS-MAA Survey (second report). *Notices of the American Mathematical Society* 43(8), 848-858.

Gaddy, C. D. (1995). Overview of current labor market conditions: Employment for new science and engineering Ph.D.s. In A. H. Teich, S. D. Nelson, and C. McEnaney (eds.), *AAAS Science and Technology Policy Yearbook*. Washington, DC: AAAS.

Ingram, L., Brown, P., and Mitchell, S. (1995). *Humanities Doctorates in the United States: 1993 Profile*. Washington, DC: National Academy Press.

Klatzky, R. L. (1990). Departments should prepare students for careers. In L. Bickman and H. Ellis (eds.), *Preparing Psychologists for the 21st Century: Proceedings of the National Conference on Graduate Education In Psychology*. Hillsdale, NJ: Lawrence Erlbaum.

Malcom, S. M., George, Y. S., and Van Horne, V. V. (1996). *The Effect of the Changing Policy Climate on Science, Mathematics, and Engineering Diversity*. Washington, DC: AAAS.

Massy, W. F., and Goldman, C. A. (1995). *The Production and Utilization of Science and Engineering Doctorates in the United States*. Stanford, CA: Stanford Institute for Higher Education Research.

Maxfield, B. D. (1982). *Humanities Doctorates in the United States: 1981 Profile*. Washington, DC: National Academy Press.

Maxfield, B. D. (1980). *Science, Engineering and Humanities Doctorates in the United*

States: 1979 Profile. Washington, DC: National Academy of Sciences.

Maxfield, B. D., Ahern, N. C., and Spisak, A. W. (1978). *Science, Engineering and Humanities Doctorates in the United States: 1977 Profile.* Washington, DC: National Academy of Sciences.

Maxfield, B. D., and Brown, P. (1986). *Humanities Doctorates in the United States: 1985 Profile.* Washington, DC: National Academy Press.

Maxfield, B. D., and Brown, P. (1985). *Humanities Doctorates in the United States: 1983 profile.* Washington, DC: National Academy Press.

National Science Board (NSB). (1996). *Science and Engineering Indicators 1996* (NSB 96-21). Washington, DC: Author.

National Science Foundation (NSF). (1996a). *Selected Data on Science and Engineering Doctorate Awards: 1995* (NSF 96-303). Arlington, VA: Author.

National Science Foundation (NSF). (1996b). *Characteristics of Doctoral Scientists and Engineers in the United States: 1993* (NSF 96-302). Arlington, VA: Author.

National Science Foundation (NSF). (1996c). *Women, Minorities, and Persons With Disabilities In Science and Engineering: 1996* (NSF 96-311). Arlington, VA: Author.

National Science Foundation (NSF). (1994). *Characteristics of Doctoral Scientists and Engineers in the United States: 1991* (NSF 94-307). Arlington, VA: Author.

National Science Foundation (NSF). (1991). *Characteristics of Doctoral Scientists and Engineers in the United States: 1989* (NSF 91-317). Arlington, VA: Author. (Contains selected time series data from 1973-1989)

National Science Foundation. (1990). *The State of Academic Science and Engineering.* Arlington, VA: Author.

National Science Foundation (NSF). (1989, April 25, working draft). *Future Scarcities of Scientists and Engineers: Problems and Solutions.* Washington, DC: Author.

Ralston, A. (1996, March). The demographics of candidates for faculty positions in computer science. *Communications of the ACM* 39(3): 78-84.

Richards, J. M., and Gottfredson, G. D. (1984). Patterns of accomplishments among psychologists. *American Psychologist* 39: 1352-1356.

Simmons, R. O., and Thurgood, D. H. (1995). *Summary Report 1994: Doctorate Recipients from United States Universities.* Washington, DC: National Academy Press.

Sonnert, G. (1995-1996, Winter). Gender equity in science: Still an elusive goal. *Issues in Science and Technology:* 53-58.

Sonnert, G., and Holton, G. (1996). Career patterns of women and men in the sciences. *American Scientist* 84: 63-71.

Tobias, S., Chubin, D., and Aylesworth, K. (1995). *Rethinking Science as a Career: Perceptions and Realities in the Physical Sciences.* Tucson, AZ: Research Corporation.

U.S. Department of Labor. (1996, October). Table 4: Employment status of the population. *Monthly Labor Review* 119(10): 99.

Wicherski, M. and Kohout, J. (1995). *The 1993 Doctorate Employment Survey.* Washington, DC: American Psychological Association.

Author Index

Subject Index

T

U

V

W

Y

Contents of Previous Volumes

VOLUME IV

VOLUME V

VOLUME VII

1991: 492 pages ISBN 0-87586-097-4

VOLUME VIII

VOLUME IX

VOLUME 10

1994: 454 pages ISBN 0-87586-111-3

VOLUME XI

1996: 464 pages ISBN 0-87586-115-6

VOLUME XII

1997: 498 pages Paper: ISBN 0-87586-118-0 Cloth: ISBN 0-87586-119-9