

# QUALITY AND ACCOUNTABILITY IN HIGHER EDUCATION

---

Improving Policy,  
Enhancing Performance

E. GRADY BOGUE  
and  
KIMBERELY BINGHAM HALL

---

PRAEGER

Westport, Connecticut  
London

## Library of Congress Cataloging-in-Publication Data

Bogue, E. Grady (Ernest Grady), 1935–

Quality and accountability in higher education : improving policy, enhancing performance / E. Grady Bogue and Kimberely Bingham Hall.

p. cm.

Includes bibliographical references and index.

ISBN 0–89789–883–4 (alk. paper)

1. Educational accountability—United States. 2. Education, Higher—United States—Evaluation. I. Hall, Kimberely Bingham. II. Title

LB2806.22.B637 2003

378.1—dc21 2002028760

British Library Cataloguing in Publication data is available.

Copyright © 2003 by E. Grady Bogue and Kimberely Bingham Hall

All rights reserved. No portion of this book may be reproduced, by any process or technique, without the express written consent of the publisher.

Library of Congress Catalog Card Number: 2002028760

ISBN 0–89789–883–4

First published in 2003

Praeger Publishers, 88 Post Road West, Westport, CT 06881

An imprint of Greenwood Publishing Group, Inc.

[www.praeger.com](http://www.praeger.com)

Printed in the United States of America



The paper used in this book complies with the Permanent Paper Standard issued by the National Information Standards Organization (Z39.48–1984).

10 9 8 7 6 5 4 3 2 1

### Copyright Acknowledgment

The authors and publisher gratefully acknowledge permission to reprint W. Edwards Deming's "14 Points for Management" (pp. 165–66, this volume) from W. Edwards Deming, *Out of the Crisis* (Cambridge: MIT Press, 1986), pp. 23–24.

44. Southern Association of Colleges and Schools. (2000). Memorandum to Member Institutions of the Commission on Colleges and Other Interested Parties, December 5.
45. North Central Association. (2001). Available on-line: <http://www.nca.org>.
46. Council for Higher Education Accreditation. (2000). "Distance Learning in Higher Education." CHEA Update No. 3.
47. *Ibid.*, 5.
48. Council for Higher Education Accreditation. (2000). "Distance Learning: Academic and Political Challenges for Higher Education." CHEA Monograph Series, Number 1, 11.
49. Oblinger, D.G., Barone, C. A., and Hawkins, B. L. (2001). "Distributed Education and Its Challenges: An Overview." American Council on Education Center for Policy Analysis, 33.
50. *Ibid.*, 35.
51. *Ibid.*
52. American Academy for Liberal Education. (2000). "Accrediting for Educational Effectiveness in Distance Learning Degree Programs." Project Proposal, 1.
53. *Ibid.*, 7.
54. Young, Chambers, Kells, and Associates, *Understanding Accreditation*, 380.
55. Selden, *Accreditation*, 32–34.
56. Mayhew, L., Ford, P., and Hubbard, D. (1990). *The Quest for Quality: The Challenge for Undergraduate Education in the 1990s*. San Francisco: Jossey-Bass, 222–224.
57. *Ibid.*, 231.
58. Greenberg, M. (2000). "America's Colleges Should Rank Themselves." *Chronicle of Higher Education*, July 16.
59. McMurtrie, B. (2001). "Regional Accreditors Punish Colleges Rarely and Inconsistently." *Chronicle of Higher Education*. January 12.
60. *Ibid.*
61. Newsom, W., and Hayes, C. (1990–91). "Are Mission Statements Worthwhile?" *Planning for Higher Education*, Winter, 19, 28.
62. *Ibid.*, 29.
63. Delucchi, M. (1997). "Liberal Arts Colleges and the Myth of Uniqueness." *Journal of Higher Education*, May/June.
64. Young, Chambers, Kells, and Associates, *Understanding Accreditation*, 13.

## Chapter 3

### College Rankings and Ratings

---

#### The Test of Reputation

In the fall of 1983, *U.S. News and World Report* first brought to newsstands across the nation a report and rating of "America's Best Colleges." Although there have been a number of scholarly ratings published since 1911, the *U.S. News'* journalistic exercises, which are now published annually, may have become the nation's most widely known quality reports. They have certainly become the most intensely debated approach to quality assurance among college educators.

After describing a range of quality indicators, McGuire and others concluded that "given a universe of unsatisfactory output concepts, reputation—with all its flaws—is probably as good or better because it is broader and more representative of the range of important output components that are produced."<sup>1</sup> In contrast, Robert Zemsky and William Massey offered this comment on the fall 1989 *U.S. News and World Report* ratings: "The rankings of colleges and universities in *U.S. News* are as meaningless as the calculations spewed by Dustin Hoffman's character in 'Rain Man'."<sup>2</sup> Graham and Thompson (2001) agree with this assessment and elaborate that *U.S. News* pays little attention to measures of learning and effective educational practices, but enormous attention to "a school's wealth, reputation, and the achievement of the high school students it admits."<sup>3</sup> Those within academe often criticize and denounce published ratings, but are quick to advertise their inclusion. The News Bureau of the University of Illinois at Urbana-Champaign touted its 1997 eighteenth ranking in the nation in *Money* magazine by stating that it was by far the highest ranked Illinois and Big Ten institution. Its news release even listed the rankings of other schools in Illinois and the Big Ten to illustrate the superior ranking.<sup>4</sup>

The entire process of rating and ranking is built on the philosophic premise discussed in our opening chapter, that quality is in limited supply. In other words, there can never be more than ten in the top ten or twenty-five in the top twenty-five, no matter that the actual performance level of a program or an institution or what impact the program or institution has on its students. Whether there is a definitive relationship between reputation and results is yet another issue to be engaged in this chapter.

Most organizations are concerned with both the perception and the reality of quality—that is, the public perception of and the hard data on product performance. One is not surprised then to see full-page advertisements in the *Wall Street Journal*, *USA Today*, or weekly magazines such as *Time* or *U.S. News and World Report* in which auto or airline industry executives pitch the quality theme. Indeed, in a 1988 *Time* magazine advertisement, Chrysler executive Lee Iacocca outlined a “quality bill of rights” to which customers of Chrysler are supposedly entitled.<sup>5</sup> Ford Motor Company has advertised today that “Quality Is Job One.” Is there a lesson here for American higher education? It may not be enough for quality to show through in whatever indicator or evidence we may select to demonstrate quality. It may be equally important for us to attend to public perceptions of quality as well.

Do rating and ranking exercises make a meaningful contribution to quality assurance? We begin the engagement of that question by exploring first some of the more significant reports on both graduate and undergraduate rankings.

### RATING GRADUATE PROGRAMS—THE ORIGINS OF COLLEGE RANKING

In his 1964 book *The Academic Man*, Logan Wilson said that rating colleges has a history as early as 1911, when the Bureau of Education published a list, a rating of 344 institutions by the Association of American Universities. The leading institutions were, in order, Harvard, Chicago, Columbia, California, Yale, Michigan, Cornell, Princeton, Johns Hopkins, Wisconsin, and Minnesota.<sup>6</sup>

The study most often cited as the original study of graduate quality was reported by Raymond Hughes in 1925. As rationale for this early rating of graduate programs and schools, Hughes offered the opinion: “It has seemed that such a rating would be of distinct value to the college president or dean who is seeking men to fill vacancies on this staff. Such a rating also seems proper and desirable in printed form, so that any one interested can turn to it readily for a rough estimate of the work in a given field.”<sup>7</sup>

The ratings began by securing from members of the Miami University (Ohio) faculty “a list of the universities which conceivably might be doing high grade work leading to a doctor’s degree.”<sup>8</sup> The Miami faculty also furnished a “list of from forty to sixty men who were teaching his subject in colleges and universities in this country, at least half of the names on the list to be those of profes-

sors in colleges rather than universities.”<sup>9</sup> Rating forms were returned, according to Hughes’s report, from about half the respondents.

Hughes chaired a second study of graduate schools for the American Council on Education, and the results were reported in the April 1934 issue of the *Educational Record*.<sup>10</sup> Whereas the 1925 study included a rating of about twenty fields, this one encompassed fifty different fields of study. Identification of the institutions offering different graduate work was derived from a study of catalogues and reports of graduate deans. Learned societies in each of the fields were asked to supply a list of 100 scholars in that field, to whom the rating forms were circulated.

At the end of his introduction to the results, Hughes commented: “There was marked evidence of a lag in the estimate of departments. A department which has been strong, but which has lost good men and is really on the decline, has in several cases been rated too high. On the other hand, several departments that have recently developed much strength seem to be underrated. If this type of study could be repeated every few years, such errors would be corrected.”<sup>11</sup> The time-lag liability of ratings and rankings is thus well identified. However, the next comprehensive rating was not to occur within the time line suggested by Hughes.

In 1959 Hayward Keniston reported a ranking of graduate departments and institutions undertaken as part of an evaluation of the program at the University of Pennsylvania. Keniston describes the study’s approaches as follows: “A letter was addressed to the chairmen of departments in each of twenty-five leading universities of the country. The list was compiled on the basis of (1) membership in the Association of American Universities, (2) number of Ph.D.’s awarded in recent years, and (3) geographical distribution. The list did not include technical schools like Massachusetts Institute of Technology and the California Institute of Technology, nor state colleges, like Iowa State, Michigan State or Penn State, since the purpose was to compare institutions which offered the doctorate in a wide variety of fields.”<sup>12</sup>

These rather important caveats notwithstanding, a review of the results will prove of interest. Appearing in the Keniston list of top twenty institutions are Harvard, California, Columbia, Yale, Michigan, Chicago, Princeton, Wisconsin, Cornell, Illinois, Pennsylvania, Minnesota, Stanford, University of California, Los Angeles (UCLA), Indiana, Johns Hopkins, Northwestern, Ohio State, New York University (NYU), and Washington. With the exception of UCLA, NYU, and Washington, all remaining institutions had appeared in Hughes’s 1925 ranking. There had been, then, a considerable stability over the thirty-two-year history of the ratings. North Carolina and Iowa were not in the Keniston list, although they appeared in 1925. Can it be that the actual quality of these two universities declined during this thirty-two-year period? We will have more to say about this feature of collegiate rankings later in the chapter.

Two of the most frequently cited and best known of recent quality ratings of graduate departments and institutions are the 1966 study by Allan Cartter and

the 1970 study by Roose and Anderson. Logan Wilson's foreword in Allan Cartter's report contains an arresting opening line: "Excellence, by definition, is a state only the few rather than the many can attain."<sup>13</sup> With this definition, Wilson supports the philosophic position of "limited supply" to which we earlier referred, which assumes a standard of performance that is comparative and relative rather than criterion based. The Cartter study is cited as the fourth such quality rating following Hughes's study of 1925, the American Council of Education Study of 1934, and the Keniston study of 1957.

Cartter offers these introductory notes on the concept of quality: "Quality is an elusive attribute, not easily submitted to measurement . . . In an operational sense, quality is someone's subjective assessment, for there is no way of objectively measuring what is in essence an attribute of value."<sup>14</sup> Arguing for the contributions of rankings and ratings, Cartter points to the limitations of diversity. "Diversity can be a costly luxury if it is accompanied by ignorance. Our present system works fairly well because most students, parents, and prospective employers know that a bachelor's degree from Harvard, Stanford, Swathmore, or Reed is ordinarily a better indication of ability and accomplishment than a bachelor's degree from Melrose A&M or Siwash College."<sup>15</sup> As an editorial note, we find it interesting that Cartter had no difficulty naming top institutions but fudged on his public disclosure of the supposedly lower-quality schools by resorting to pseudonyms and poorly disguised racial identification. Cartter then states this rationale for quality ratings: "Just as consumer knowledge and honest advertisement are requisite if a competitive economy is to work satisfactorily, so an improved knowledge of opportunities and of quality is desirable if a diverse educational system is to work effectively."<sup>16</sup>

To avoid criticisms of the earlier studies, Cartter used three panels of raters—department chairpersons, senior scholars, and junior scholars—in twenty-nine different fields of study.

In a replication of the Cartter study in 1970, Roose and Anderson published yet another rating of graduate programs. In a synopsis of the findings, they noted: "There is an increase in the rated quality of faculty of graduate programs, moving from 69.8 percent rated adequate plus or better in 1964 to 80 percent in 1969. Second, there is evidence of regional improvement. The south, for example, had only 59 percent of its faculties rated 'adequate plus' or better in 1964 but this percentage rose to 73 percent in 1969."<sup>17</sup>

In a section entitled "policy implications of the findings," Roose and Anderson warn of the limitations: "The superficiality of exclusive reliance on reputation as a measure of quality was well illustrated in a comment made to us by the former chief academic officer of a top rated institution about one of his distinguished faculty members: '\_\_\_\_\_ is highly regarded in his profession and has contributed importantly to the reputation of his department, but what has he done for students? In twelve years he has not turned out a single Ph.D.'"<sup>18</sup>

While we are reviewing the Cartter and Roose-Anderson reports, we note one of the more vigorous assaults on these two reports. In a monograph entitled *The*

*Ranking Game: The Power of the Academic Elite* by W. Patrick Dolan (1976), we were drawn not so much to the body of the monograph as to the preface. William A. Arrowsmith begins this preface by labeling the reports as "quantified gossip" and "formidable professional sanctions against daring, diversity, and openness." Arrowsmith's prefatory remarks include the following judgement: "The Cartter report and the Roose-Anderson reports are monstrosities, not simply because they are the patent product of reflex, bad faith and suboptimization, but because they are the regressive instruments of standardization which has no valid cultural or human purpose; which serves merely governmental, professional, administrative, or bureaucratic convenience. They simply have no educational purpose which is compatible with the needs of our culture or with the life of the mind."<sup>19</sup>

Of this we may be certain: The person who dares to evaluate anything in higher education invites evaluation. We will never be without lively argument when it comes to the purpose and performance of American collegiate education.

In 1976, Mary Jo Clark, Rodney Hartnett, and Leonard Baird of the Educational Testing Service (ETS) reported on an extensive study of quality indicators in three graduate fields. The opening pages of their report carry these warnings about the contributions of program rating and give a rationale for multidimensional study of graduate quality: "Ratings of the reputation of a program among faculty members in the same field have a place in program evaluation; but they are not very helpful to those who may be seeking to improve their program, are highly related to program size and visibility, and only occasionally reflect recent changes (good or bad) in a program."<sup>20</sup> Their study involved an extensive review of the variety of evidences that should be examined to ascertain program quality: faculty, students, resources, and curricular characteristics.

As another example of graduate study ranking, we explore a five-volume study that attempts to take advantage of the multidimensional approach advocated by the previous ETS study by Clark and others. The 1982 study edited by Jones, Lindzey, and Coggeshall, conducted under the aegis of the Conference Board of Associated Research Councils and published by the National Academy of Sciences, is by far the most ambitious of contemporary efforts—reviewing programs at 228 institutions (compared, for example, to 130 in the Roose-Anderson study). Within these 228 institutions, almost 2,700 programs in thirty-two disciplines are evaluated.

In their preface to Volume 1, the editors noted that at the time of the study American universities were producing over 20,000 research doctorates a year but "what might surprise us, however, is the imbalance between the putative national importance of research-doctoral programs and the amount of sustained evaluative attention they themselves receive."<sup>21</sup> They claim that collegiate educators are poorly informed about the quality of graduate programs and offer their study as a logical but improved continuance of the Roose-Anderson tradition.

This study employed sixteen measures for most of the fields reviewed: mathematical and physical sciences, humanities, biological sciences, engineering, and social and behavioral sciences. These sixteen variables were clustered in six groups: program size (three measures), characteristics of graduates (four measures), reputational survey results (four measures), university library size (one measure), research support (two measures) and publication records (two measures). Many of the measures are of dubious qualitative linkage. The number of faculty members in a program (program size measures) and the median number of years from first enrollment of a doctoral student to receipt of the doctorate (characteristics of graduate measures) may be of interest but hardly qualify as hard evidence on quality.

The qualitative heart of this study is embraced by the four "reputational survey results," in which participating faculty were asked to evaluate these factors: the scholarly quality of program faculty on a scale from 0 to 5, the effectiveness of the program in educating research scholars, and the improvement in program quality in the last five years. Then the faculty evaluators were asked to indicate their familiarity with the programs they were evaluating. In Volume 1 of the evaluation of programs in mathematics and physical sciences, the comments of the editors are instructive: "It should be pointed out that the evaluators, on the average, were unfamiliar with almost one-third of the programs they were asked to consider."<sup>22</sup> Is this a condition designed to enhance confidence in the outcomes of the evaluation? We think not.

Under the measure of publication records, the editors report that a program was evaluated on the number of published articles attributed to it. The qualitative limits of this measure are reasonably obvious, but it ignores other major evidences of scholarship and also tends to bias the evaluation toward large departments.

In the May/June 1983 issue of *Change*, David Webster furnishes an informing review of the 1982 study. He is both complimentary and critical and describes the study as "the biggest, best, most expensive, most thoughtfully conceived, and carefully carried out academic quality rating ever done."<sup>23</sup> Unlike what one might expect from a study of this nature, one searches in vain for systematic rankings of programs based on any one of the sixteen measures or combinations of measures. Webster criticizes the authors for this approach: "By publishing endless columns of figures for institutions in each discipline it covers, without summarizing, combining, or averaging these figures, or ranking the institutions in any way, this most reluctant of all academic quality ratings is like the bible for some religious sects. Of which anyone is welcome to make any interpretation he or she wishes."<sup>24</sup>

Webster then undertakes to remedy this difficulty by a numeric routine designed to portray shifts in institutional rankings from the earlier rankings (by Hughes, Keniston, and so on) to the Jones study. He publishes a table of results showing, among other movements, that Johns Hopkins had slipped twenty-three positions and MIT had gained fourteen positions from the 1925 Hughes study. Webster admits, however, that the results of this comparison cannot be consid-

ered very accurate for the simple reason that different groups of judges used different criteria with different groups of departments and different groups of universities.

Returning now to the original 1982 study, we find on page vi of the preface to Volume 1 that the three purposes of this massive study were to assist students in seeking the best match of graduate programs to their interests, to serve scholars in higher education, and to inform the judgement of those responsible for protecting quality. As an interesting and perhaps questionable commentary on these purposes, the copy of the study that we reviewed had never been checked out of the research university library in which we found it.

Webster cites this 1982 study as one of the more expensive ones undertaken. Given the general liabilities of previous studies and the specific liabilities of this 1982 study, we wonder if funding agencies ever bother to obtain hard evidence on whether the three purposes above were well served and whether the study had any significant decision utility. It would be a simple matter, for example, to evaluate the achievement of the first purpose just cited by simply asking a random sample of students enrolled in research-doctoral programs whether they had ever heard of the study.

In 1995 the National Research Council (NRC) released *An Assessment of Research-Doctorate Programs in the United States*. According to Webster and Skinner (1996), the NRC release covers 41 disciplines, 274 institutions, 3,634 programs, and the reputational judgements of 7,876 faculty members.<sup>25</sup> Not only does this rating of programs include reputational rankings by faculty but also objective statistics relating to the seniority, research productivity, and productivity in conferring doctoral degrees of program faculty.<sup>26</sup> "Institutions were ranked if they offered even one program from among the 41 disciplines the Report covered that met the Committee's eligibility criteria."<sup>27</sup> Top institutions in the arts and humanities included the University of California, Berkeley (receiving a rating of 4.36 on a 5.0 scale) and Princeton University (rating of 4.28). The following categories also yielded other first-place recipients: biological sciences—the University of California, San Francisco (4.67), engineering—the Massachusetts Institute of Technology (4.65), physical sciences and mathematics—the University of California, Berkeley (4.74), and social and behavioral sciences—Harvard University (4.61). According to Rogers and Rogers (1997), this publication has been criticized because it does not include master's and doctoral-level programs in fields such as law, medicine, business, dentistry, nursing, social work, architecture, library science, journalism, education, and public administration. In addition, NRC only conducts its study every ten years and it is very expensive. This time frame limits the usefulness of the ratings.<sup>28</sup>

The Roose and Anderson or NRC studies did not involve, by the way, major rankings of professional programs, other than engineering. Perhaps we should examine some of the work that has been done in such rankings. In 1973 Margulies and Blau ranked professional schools in such fields as medicine, law, and so forth. The authors sent questionnaires to deans in 1,180 schools offering de-

gress in seventeen different professional program fields. A comment on response rate in this study is instructive: "A look at the response rates shows that one explanation for the differences (in response rates) between types of schools is the prestige of the profession. The largest proportions of deans who did not respond are in those professions that research has shown to have the highest prestige. Inasmuch as the ratings of expert judges are of interest, even if they are statistically unreliable, we present the rating of the eight types of professional schools with fewer than 20 respondents in the lower part of the table."<sup>29</sup>

The authors seem to suggest that we should be interested in ratings of these programs even though the representative participation of raters cannot be assured. It is no small curiosity that we are invited to find these results of interest even though they are statistically meaningless, and professionals in these high-prestige fields did not care enough to participate.

A second example is the *U.S. News and World Report* rating of graduate and professional programs. Not content with their venture into the rating of colleges and universities, *U.S. News and World Report* entered the field of graduate school rating in March 1990. This issue offered ratings in four graduate and professional fields: business, law, medicine, and engineering.<sup>30</sup> *U.S. News and World Report* now publishes rankings for more than 1,000 graduate programs surveyed each year in the five primary fields of law, education, business, engineering, and medicine. However, the issue is of interest for more than the ratings. With the publication of the first graduate schools rankings is an accompanying article entitled "Why U.S. News Ranks Colleges" and it opens with this lead: "The sad truth is that it is easier to learn about the relative merits of compact disc players that it is to compare and contrast America's professional schools. And some educators prefer to keep it that way."<sup>31</sup>

Here is the closing comment of *U.S. News* editors: "The editors expect that like our study of undergraduate education, now going into its sixth edition, it will evolve, change and eventually merit the active cooperation of all the leaders of the graduate and professional community. We trust they will come to understand that the 'light of public scrutiny' often can be as valuable a tool for improving higher education as it is for informing the reading public."<sup>32</sup>

Discomforting though it may be to some collegiate educators, *U.S. News and World Report* makes a good point. We cannot continue to talk about quality without operational evidence of what we are talking about, and that is going to mean public disclosure. It is difficult to reduce the quality of an institution to a single number, but what other methods have been developed? We are inclined to see the test of public disclosure—whether on ratings, outcomes assessments, results of peer reviews, or other evidence—to be a helpful trend. We encourage readers to review the rating results of graduate and professional schools in the five fields cited: business, education, law, medicine, and engineering.

In summary, what can we say about the reputational ratings and rankings of graduate programs and institutions having graduate programs? We certainly have to note an important consistency and stability in the results. One is hard

pressed to find any list of top-rated institutions in which some of the national universities of the land cannot be found. The visibility and prestige of these universities, and the primary basis of these ratings, rest heavily upon the eminence of their faculties; and the eminence of the faculties resides heavily in their publication and research records. We are not surprised, then, that the top-rated universities and graduate programs are those whose faculties are publication productive and research oriented. Institutional size, history, and resources play important roles in this stability.

### REPUTATIONAL RATINGS OF UNDERGRADUATE PROGRAMS

Although the rating of graduate programs in the United States has a history reaching back to the early part of the twentieth century, the rating of undergraduate programs is of more recent origin. Of special interest is that a good portion of activity and initiative for such ratings has occurred not so much in scholarly studies, such as those previously discussed in this chapter, but in media and journalistic reviews.

A good lead example for the ratings of undergraduate programs is furnished by Lewis Solmon and Alexander Astin, who conducted a pilot study to rate undergraduate programs in seven fields in four states—California, Illinois, North Carolina, and New York. A more complete record of their study can be found in the September and October 1981 issues of *Change* magazine.<sup>33</sup>

Respondents were all department members, approximately 1,000 in number. Rating criteria included these factors:

- Overall quality of undergraduate education
- Preparation of student for graduate school
- Preparation of student for employment
- Faculty commitment to undergraduate teaching
- Scholarly accomplishments of faculty
- Innovativeness of curriculum

Here is one interesting, and perhaps disturbing, result of the study. Princeton University appears in the top ten departments of business on each of the six criteria above even though Princeton does not offer an undergraduate program in business. Solmon and Astin comment as follows: "The finding underscores the need to interpret the results of any reputational survey with care, since strong halo effects appear to be operating. In this report, we have not presented the results for business programs because of the confounding halo effects."<sup>34</sup>

In the follow-up article entitled "Are Reputational Ratings Needed to Measure Quality?" Solmon and Astin conclude: "In short, while our analysis sug-

gests that reputational ratings of undergraduate programs may indeed be unnecessary because they seem to be redundant with other known information about institutions, we must defer final judgment about the value of such ratings until additional ratings covering more fields and possibly more diverse quality criteria can be obtained, and until longitudinal value-added studies can be carried out to test the validity of such ratings.<sup>35</sup> The last comment merits more thought because it offers a good lead-in to chapter 7 on student outcomes assessment. Does attending a highly rated institution result in a demonstrated impact on student learning?

Other dimensions of the Solmon and Astin study also deserve attention. In their opening commentary, the authors point out that many raters chose not to respond for a variety of reasons: burdensome time commitments, lack of knowledge about departments to be rated, confusion over institutions included in the lists, and so on. The response rate is not cited in the article.

Another provocative finding of the Solmon and Astin study was that the position of an institution on the two lists of institutions provided to raters—a national list and a state list—did affect how raters responded. Indeed, the results were in opposite directions in two different fields. “California sociologists tended to rate the ‘scholarly or professional accomplishments of faculty’ in California institutions lower when those institutions were included in the national list rather than the California list. Chemistry faculty in New York tend to rate chemistry departments in their own state more favorably when they are included on the national list rather than a separate list, whereas sociology faculty in New York tend to rate their department more poorly when they are included in a single list along with the other national institutions.”<sup>36</sup>

What can we say about this study of ratings in undergraduate departments? The simplest answer may be that if one knows something about the admissions selectivity and size of a campus, then one can reasonably expect that institution to be viewed as a high-quality institution. Second, a variety of variables can affect the reliability and validity of these rankings—size of response, composition and arrangement of institutional lists, and so on. And finally, one cannot know from reputational rankings what direct educational benefit is conferred on a student; that is, whether the institution does, in fact, make a value-added difference. Our common sense is rewarded in this case. Many have experienced graduates of highly prestigious schools whose minds seem empty of meaning and hearts empty of caring. And many have seen graduates from “Melrose A & M and Siwash College”—to borrow pseudonyms from Cartter—whose minds and hearts are making a difference.

As stated earlier, easily the most visible and widely known rating activities related to undergraduate programs in recent years are the surveys conducted by *U.S. News and World Report*. In 1983, 1,308 presidents were asked to name the nation’s highest-quality undergraduate schools. Fifty percent responded. Commenting on the fact that only five of the seventy-six schools mentioned were public institutions, *U.S. News and World Report* writers observed: “Educators

point to conditions found at many taxpayer supported institutions—among them larger class sizes, more graduate students serving as instructors for undergraduates and less selective admissions standards based on serving state residents than on attracting the nation’s top students—as reasons for the predominance of private schools on the lists.”<sup>37</sup>

Four years later, in the 1987 survey, 16 of 144 schools would be public, an increase from 6.5 percent to 11.1 percent of public institutions in the ratings. Can we suppose that quality conditions in public colleges changed that dramatically in four years?

Included in the 1985 survey were 1,318 institutions that offered liberal arts programs, with professional schools and military academies excluded. As in 1983 and 1987, the raters were college presidents, each of whom was asked to select the top five undergraduate schools from a list of colleges and universities similar to his or her campus in mission. Of the 1,318 presidents surveyed, 788 or 60 percent responded.

One of the best features of the *U.S. News and World Report* 1985 rating is the focus on diversity and innovation. For example, Evergreen State College in Washington was cited for its innovative curriculum in which students study special topical courses in small student-faculty settings. Evaluation of students at Evergreen takes the form of narrative entries rather than grades. Alverno College in Milwaukee, a smaller Catholic school for women, was cited for its development and evaluation of general education competencies in communication, analysis, problem solving, valuing, social integration, responsibility for global environment, responsible citizenship, and esthetic responsiveness. And Trinity University, a private liberal arts school in San Antonio, Texas, was cited for its aggressiveness in fund-raising and the attraction of top scholars with endowed chairs.<sup>38</sup>

For the 1987 survey, 1,329 presidents were asked to rate colleges with missions similar to their own, as in the 1985 survey. Of the 1,329 polled, 650 responded—a 60 percent response ratio, as in 1985. Authors of the article indicate that “a small number of the presidents declined to participate in the survey and wrote to the magazine to say they felt that neither they nor their peers were in a position to judge the academic quality of institutions other than their own.”<sup>39</sup> For 1987, the presidents were asked to rate the top ten institutions, rather than the top five as in the 1985 survey.

Once again, the writers of the story emphasize the curricular innovations that draw attention to the diversity of American higher education: the great books curriculum at St. John’s, global education emphasis at Earlham, value-added assessment at Northeast Missouri State University, freshman humanities seminars at Wofford College in South Carolina, one-course-at-a-time at Colorado College.

An informative addendum to the 1987 study is entitled “How the Rankings Changed: ‘The Ups and Downs’.” It begins with this lead-in statement: “It doesn’t take long for new college presidents to discover an essential truth about



their vocation: Academic reputations are like campus trees, slow to grow, but once established slow to die."<sup>40</sup>

However, a number of schools dropped out of the 1987 ratings, and the authors felt an explanation was in order. They explain that the disappearance of schools like Mills College in California, Wheaton in Massachusetts, Austin College in Texas, Old Dominion in Virginia, and Tennessee Tech came about because they were victims of their own success. These institutions and others had moved into more competitive Carnegie classifications.

Explaining why Harvey Mudd College, which ranked first among small comprehensive institutions in the 1985 survey, dropped completely out of the rankings, the writers said: "Harvey Mudd's slide into seeming academic oblivion is more apparent than real. Earlier this year, Carnegie reclassified Harvey Mudd, which emphasized engineering, chemistry, math, and physics courses, as a specialized institution of engineering and technology. Because the *U.S. News and World Report* survey does not include a category for ranking specialized schools, Harvey Mudd did not even appear on this year's ballot. In brief after 1985, there was change in Harvey Mudd's category, not its quality."<sup>41</sup> (This conclusion, both descriptive and evaluative, we should remember, is not the consensus of presidential raters but of the authors.)

In October 1988, *U.S. News and World Report* departed from the practices of its three previous surveys, revealing what was done and why: "This year, however, after extensive consultation with college presidents and other academic experts, *U.S. News and World Report* has made two major changes in its study. First, because academic deans and admissions officers often see education from rather different perspectives than do college presidents, they also have been included in the survey of more than 1,000 college officials. Second, because the expert opinions are just that, opinions, *U.S. News and World Report* has based its largest academic rankings on objective data as well as on the subjective judgments in the survey."<sup>42</sup> The authors then indicated that these data included five measures of quality: admissions selectivity, faculty strength and instructional budget per student, resources for educational programs, graduation rates, and school reputational studies.

What happens to the reputational quality of a college or university when one changes both the rater and the criteria? Table 3.1 contains the rankings for national universities in the 1987, 1988, 1989, and 1990 surveys. In 1988, four state institutions dropped out of the top twenty-five: the University of Wisconsin, the University of Illinois, the University of Texas, and the College of William and Mary. And how can we explain the wide swings in position for an institution such as the University of California, Berkeley, which in successive years occupied rank positions of five, twenty-four, thirteen, and thirteen among the national universities?

Did the real quality of these schools change that much in one year? Perhaps they would have liked to return to the more "global and subjective ratings" of college presidents rather than the more "objective ratings" rendered by deans

Table 3.1  
*Reputational Ranking of National Universities*  
by U.S. News and World Report

1987	Rank			School (State)
	1988	1989	1990	
3	1	1	3	Yale University (Conn.)
4	2	2	4	Princeton University (N.J.)
21	3	4	5	California Institute of Technology
2	4	3	1	Harvard University (Mass.)
11	5	7	6	Massachusetts Institute of Technology
1	6	6	2	Stanford University (Calif.)
6	7	8	8	Dartmouth College (N.H.)
18	8	12	10	Columbia University (N.Y.)
14	9	10	16	Rice University (Tex.)
8	10	9	11	University of Chicago (Ill.)
16	11	14	15	Johns Hopkins University (Md.)
7	12	5	7	Duke University (N.C.)
10	13	15	12	Brown University (R.I.)
11	14	11	9	Cornell University (N.Y.)
19	15	20	13	University of Pennsylvania
17	16	19	23	Northwestern University (Ill.)
*	17	25	19	Georgetown University (D.C.)
*	18	23	*	University of Notre Dame (Ind.)
23	19	22	24	Washington University (Mo.)
15	20	21	18	University of Virginia
*	21	16	17	University of California, Los Angeles
25	22	*	*	Emory University (Ga.)
11	23	18	20	University of North Carolina, Chapel Hill
5	24	13	13	University of California, Berkeley
8	25	17	21	University of Michigan, Ann Arbor
23	*	*	*	University of Wisconsin
20	*	*	*	University of Illinois
25	*	*	*	University of Texas
22	*	*	*	College of William and Mary (Va.)
*	*	24	*	Vanderbilt University (Tenn.)
*	*	*	22	Carnegie Mellon (Pa.)
*	*	*	25	University of Rochester (N.Y.)

\*Not ranked this year.

Source: "America's Best Colleges," *U.S. News and World Report*, 1987, pp. 49-87; 1988, pp. C3-C32; 1989, pp. 53-82; 1990, pp. 103-134.

and admissions officers. At the very least, four universities felt the frustration of other schools that had not appeared in any of the *U.S. News and World Report* rankings, and perhaps they too were scratching their heads about the meaning and value of reputational studies of quality.

Modifications to the methodology used for determining the rankings of colleges and universities continued in the ten years to follow. In 1996, rankings of approximately 4,200 college presidents, deans, and admissions directors counted for 25 percent of the overall ranking total, admissions selectivity 5 percent, faculty resources 20 percent, financial resources 10 percent, retention 20 percent, and alumni giving 5 percent. A new indicator was added in this year: value added. The authors explained that this indicator counted for 5 percent of the overall ranking and accounted for the educational value added by individual colleges.

"Developed in consultation with academic experts, it focuses on the difference between a school's predicted graduation rate—based upon the median or average SAT or ACT score of its students and its educational expenditures per student—and its actual graduation rate."<sup>43</sup> By 1999, the value-added indicator was no longer used in the rankings.

In 1999, the methodology was changed once again. According to Graham and Morse (1999): "One key change we made was to employ a procedure, known as 'standardization,' that brought our calculations more into line with accepted statistical practices. In doing so, we dropped a step we had used for national universities that tended to flatten out large disparities between schools in their performance on each indicator of quality."<sup>44</sup> An example used by these authors to illustrate this change involved three universities. "School A might have spent \$200,000 per student, School B \$100,000, and School C \$50,000. If they were the leaders in spending, we would have recognized them as first, second, and third without giving them credit for the size of the disparities among them. This year our weighting system takes into account the sizes of their differences."<sup>45</sup> As a result of this change the California Institute of Technology moved into first place in this year, ahead of Harvard, Yale, and Princeton. "Caltech's spending on instruction and education-related service works out to far more per student (\$192,000) than at any other institution."<sup>46</sup> Johns Hopkins is an example of another institution that moved up in the rankings, to number seven, as a result of this methodological change.

Though thought to improve the methodology, in the September 11, 2000, issue of *U.S. News and World Report*<sup>47</sup> the authors acknowledged that the change still did not reflect how funds spent on students affected undergraduate programs—many large research institutions spend a large portion toward graduate programs. So, for the rankings published in this 2000 issue, the methodology was once again adjusted—"we adjusted each school's research spending according to the ratio of its undergraduate to graduate students."<sup>48</sup> As a consequence, Caltech and Johns Hopkins scaled back to fourth and fifteenth, respectively.

No new methodological changes were announced in the September 17, 2001, issue of *U.S. News and World Report*.<sup>49</sup> This issue included the magazine's most recent rankings of institutions. A total of 4,087 university presidents, provosts, and deans of admission participated in the survey. Princeton and Harvard Universities received top honors.

This exploration of media rankings would not be complete without reference to *Money* magazine's annual publication of "America's Best College Buys," which ranks the top 200 schools for the money. "If assessing academic quality is tough, trying to calculate a school's value—its quality relative to price—is an even more difficult task."<sup>50</sup> The magazine used statistical analysis to determine how much each school might be expected to cost, based on sixteen measures of academic performance, including entrance exam scores, class ranks and high school grades of entering students, faculty resources, quality of instruction, stu-

dent-faculty ratios, library resources, graduation rates, retention of first-year students, money spent on instruction and student services, and the academic and career success of alumni. Then, that figure is compared to actual cost. According to the September issue, the top ten schools in 2001 were:<sup>51</sup>

1. California Institute of Technology
2. Rice University
3. University of North Carolina-Chapel Hill
4. SUNY-Binghamton
5. Spelman College
6. New College of the University of Southern Florida
7. College of New Jersey
8. Truman State University (Mo.)
9. SUNY-Geneseo
10. University of Florida

*Time*, *Newsweek*, and *Business Week* also publish special issues or guides that analyze various types of educational institutions.

To this point, we have made no mention of a cluster of rating reports that can be found in most collegiate libraries: *The Gourman Reports* offers a rating of both undergraduate and graduate programs to two decimal places. "Gourman has been unofficially evaluating post-secondary educational institutions since 1955 and publishing his rankings since 1967."<sup>52</sup> In the 1983 rating of undergraduate programs in Louisiana, for example, the top three schools were Tulane, with a rating of 4.53; Louisiana State University and Agricultural and Mechanical College, with a rating of 4.35; and Louisiana State University in Shreveport with a rating of 3.38. Centenary College, which has appeared in the *U.S. News and World Report* lists of best liberal arts colleges for more than one year, was ranked thirteenth in the state with a rating of 3.06. What do these ratings mean? Not much.

*The Gourman Reports* are filled with puffed-up rhetoric about their own importance, and there are grammatical inaccuracies throughout. Nowhere in these reports are we told precisely what data were gathered, by whom they were obtained, by what means, or on what date. Nor are we told who rendered the judgments that led to the ratings expressed to the hundredths of a decimal point. Yet *The Gourman Reports* have been used by economists and other scholars studying the relationships between college quality and a host of other variables, such as alumni earnings, student choice, and so on.

In a devastating critique of *The Gourman Reports*, Webster concluded that "Jack Gourman is not a reasonable arbiter of educational quality. Students, librarians, book reviewers, and scholars need to reexamine his reports and pronouncements on college and university performance. And more importantly, we

need to look again at our willingness to accept such rankings from an individual who simply has the power of the printed page behind him."<sup>53</sup> Our fondness for numbers and our search for "number one" can lead even academics down unhappy paths.

What role do these rankings play in American higher education? We will visit that question again at the conclusion of this chapter. In terms of one of the principal purposes of quality assurance—improvement—the rankings by *Money* magazine, *U.S. News and World Report*, and *The Gourman Reports* are relatively empty. They may prove of some value to those who have the financial capacity to afford these top-ranked colleges and to those who can move from where they are to where these schools are. For those who live within service and financial range of these schools, the rankings may be of some understandable pleasure. For the schools appearing in the rankings, there is the pleasure of national recognition, perhaps accompanied by enhanced enrollment attraction, though we want to say more of that in a moment.

Returning to the discussion on quality definition of chapter 1, it may help to remember the complexity of variables associated with the definition and measurement of quality. To the customer who cannot afford a high-priced luxury automobile, its quality is irrelevant. To the student of modest means living in rural Tennessee or Texas, America's best college buy may be Columbia State Community College or Kilgore Junior College. (It is of some interest that this entire sector of American higher education—community colleges—is omitted in both the *Money* guide and the *U.S. News and World Report* rankings.)

Americans love numerical shorthand, as noted by Robert Hutchins: "He (the American) is not at home with anything he cannot count, because he is not sure of any other measure. He can not estimate or appraise quality. This leaves him with quantity."<sup>54</sup> The competitive edge in our national personality leads us to label rather than understand, to prefer averages rather than depth. There is something highly satisfying about the weekly polls that describe who is on top in collegiate football and basketball. Stories on recruiting violations, ethical misconduct of both coaches and alumni, or misuse of drugs by players will not stay our pleasure in finding out who is in the top ten or twenty.

## QUALITY RATINGS AND QUALITY REALITY

Here, then, is a brief glimpse of the history of ratings and rankings as an instrument of quality assurance. Do these ratings make a substantive contribution to quality assurance? What are the limitations? Let us begin our summary by posing a few questions suggested by a review of the studies previously cited.

*Who is an effective judge of program and institutional quality?* A variety of raters has been employed by these studies. It certainly seems important to know who is doing the rating and whether the rater has reasonable knowledge and authority on which to base his "informed opinion." The earlier graduate rankings

used department chairpersons as raters, but Allan Cartter criticized this approach, suggesting that chairpersons were not always the most distinguished scholars in their fields. Later the studies used deans and faculty in departments and programs being rated. The *U.S. News and World Report* ratings of undergraduate programs first used presidents as raters. To what extent might the judgments of these presidents be as much influenced by their knowledge of other presidents as their knowledge of the institutions in question? In few if any of the studies do we find graduates as raters, with the exception of the study by Clark, Hartnett, and Baird (1976).

*What criteria are used by the raters to judge quality?* Earlier studies obviously touched on the criteria of faculty accomplishments and library holdings. Later studies, such as those in *U.S. News and World Report*, asked for more global assessments. The study of undergraduate quality by Astin perhaps utilized the most extensive list of criteria. Astin's findings, however, lead to the next question.

*To what extent do reputational ratings contribute information not already available from other sources?* A number of these studies suggest that if one already has knowledge about program size, selectivity, library holdings, and number of graduates, one can predict program and institutional standing in the ratings.

*Is there a relation between perceived prestige, as reflected in ratings, and perceived influence?* In a study sponsored by and reported in *Change* magazine, Richard Johnson stated: "There are at least three senses in which the term 'leading' can be used: (1) prestige—the degree to which an institution is looked up to or admired; (2) innovation—the frequency with which an institution is first in generating new ideas or programs; and (3) influence—the degree to which other institutions follow the leading institutions' example. Here then is a paradox: All institutions tended to agree on a set of leading institutions which influences other colleges and universities. But when reporting on the influence on their own institutions, they also agreed that these national influential institutions do not influence them."<sup>55</sup>

*How extensive is the "halo" effect?* We recall the finding in Astin's study of undergraduate programs in which Princeton University was identified as a leading institution in undergraduate business, even though it did not have an undergraduate program in business. Do strong departments and programs tend to bathe weaker programs in a positive light? The evidence suggests that this is so.

*Is the philosophy of limited supply of excellence justified?* The use of ratings as an instrument of quality identification assumes, as we have shown, that excellence is indeed in limited supply. As additional schools develop high performance records on any criterion of quality—student growth and performance, faculty research and publication, library holdings, or curricular innovation—there can still be only five in the top five or ten in the top ten.

*What about the stability of ratings?* There appear to be mixed results on this question. One cannot argue with the apparent stability of the national univer-

sities' profile in the ratings of graduate programs. However, there are movements in and out of those ratings, suggesting that quality is a fleeting condition, and our common sense suggests otherwise. In the first *U.S. News and World Report* ratings of undergraduate institutions, only 6.5 percent of the institutions were public. Just four years later that percentage had almost doubled to 11.1 percent. Had the quality of public institutions changed that much over four years?

In this chapter, we have pointed to other equally puzzling illustrations of instability. For example, we noted the positioning in the *U.S. News and World Report* ratings for the University of California, Berkeley—from fifth place in 1987 to twenty-fourth place in 1988 and then back up to thirteenth place in 1989 and 1990. Surely the quality of this national university cannot be that mercurial in actuality. And what can we say about the May 2001 number one ranking of Dartmouth College in the business-school ratings published by the *Wall Street Journal* and the number 10 rating in the September *U.S. News and World Report* list?

*What effect does the structure of rating studies have on the response? We have evidence that the way in which questions are posed and studies structured can have an effect on the results.* Harvey Mudd College was ranked first in the 1985 *U.S. News and World Report* study but, because of a classification fluke, did not even appear in the 1987 study. Astin found that raters responded differently in their ratings depending on whether an institution appeared on a national list or a state list. In 2001, *U.S. News* recategorized many institutions due to the change in the Carnegie classification system. As a result, approximately fifty new schools were added to the rankings.

*What can we say about rate of response and confidence in ratings? We have seen that ratings are offered as a justifiable expression of quality because they constitute an aggregation of informed opinion.* Yet most of the studies report response rates of 50 to 60 percent. Although these response ratios are respectable for questionnaire research, there is little discussion on the margin of error in conclusions occasioned by these response ratios.

*Do ratings make meaningful contributions to program and quality improvement? In their informative monograph *A Question of Quality: The Higher Education Ratings Game*, Lawrence and Green (1980) point out that the "results of such studies contribute little to a program's self-knowledge or its efforts toward improvement."<sup>56</sup> Global ratings simply do not offer the specific information necessary for improvement, although program improvement is cited by most educators and evaluators as the principal reason for any venture in educational evaluation. Thus ratings must be counted as making a relatively empty contribution to this important goal.*

*Do ratings tend to encourage labeling? In their widely read book *The Academic Marketplace*, Caplow and McGee used the terms "major league, minor league, and academic Siberia" as florid terms often used to describe various sectors of American higher education.<sup>57</sup> The mixed use of metaphors from baseball*

and political geography describes an assumed pyramid of prestige topped by the national universities, which, as we have seen, everyone is willing to identify. There are no ratings, however, of the five or ten worst institutions or programs, and no one is willing to venture their names. They remain covered by Cartter's barely disguised terms "Melrose A&M and Siwash College."

*To what extent do ratings encourage a limitation of esteem as well as mission? Reputational studies have succeeded in identifying the diversity existing in American higher education.* Read again, for example, the rich descriptions of undergraduate programs found in the *U.S. News and World Report* studies. Whether these studies promote a diversity of esteem and respect for the variety of missions in American higher education is not clear.

One of the finest books ever written on the subject of excellence is Gardner's *Excellence*, first published in 1961 and revised in 1984. In perhaps the most often quoted line from that book, Gardner advocated a diversity of esteem and high expectations in these words: "The society which scorns excellence in plumbing, because plumbing is a humble activity, and tolerates shoddiness in philosophy because it is an exalted activity will have neither good plumbing nor good philosophy. Neither its pipes nor its theories will hold water."<sup>58</sup>

*Do ratings make a significant contribution to consumer choice decision? This argument has been advanced by more than one source cited in this chapter.* The essence is that a college education is a product (similar to a compact disk or an automobile), that potential students and their parents need to know something about the quality of that product, and that ratings furnish this discriminating evidence. One has to ask, however, whether potential students have been noting the wild swings in the rankings, the disappearance of some universities, and the in-and-out placement of some as carried in two different journalistic publications.

On this topic of consumer choice, we repeat an observation made earlier in this chapter. When almost 40 percent of those attending undergraduate schools in this nation attend two-year colleges, and those two-year colleges are not included in the media rankings, a legitimate question can be asked about the value of the rankings to the majority of college students in this nation. We wonder whether the variable most closely correlated with these rankings is the sales volume of the issue; if you are in the business of selling magazines, that is of legitimate and keen interest.

Thus, we have some questions concerning the decision value of ratings to potential students. Here, however, is a good topic for a useful thesis or dissertation. To plot the history of enrollments, gifts, and other benefits that supposedly follow discriminating student choice based on ratings might help bring speculation on this topic to heel. What would the facts tell us about these relationships?

*To what extent do ratings recognize the occasional "maverick" character of American higher education? One of the conventional quantitative signals of academic quality is the number or percentage of the faculty with a doctoral de-*

gree. Although few would argue with the general validity of that indicator, is there room in our reflections and ratings for the power of exceptions? Webster carries the point in this comment: "One frequently used 'objective' measure of rankings of institutions and individual departments was by the percentage of their faculty, or senior faculty, who possessed Ph.D.'s. Around 1970, one well-known department would have been ranked quite low in such a quality ranking—fully three of its senior professors had no Ph.D. One had a law degree, one had only an M.Ed., and the third had only a B.A. yet these three members of Harvard's social relations department—David Riesman, Christopher Jencks, and Erik Erikson—might possibly have been better scholars and teachers than some professors who did possess earned doctorates."<sup>59</sup>

*To what extent do ratings and rankings cause abuse of information and effect on policy?* Polocano (2001) outlined ten easy steps to a top twenty-five MBA program. The recommended steps include hiring a top public relations firm; to increasing services to recruiters including valet parking, free meals, and gift baskets; and providing a wide variety of student services for MBA students such as free parking and a fitness center. His ninth recommendation is to "Adjust your admissions policies and make up your own rules so that you can report both admissions and placement data in a more favorable light. Hire staff members to collect and massage the data and respond to media surveys in the manner you prescribe. (*U.S. News* collects self-reported data that up to this point is not audited.)."<sup>60</sup> Obviously one of the least expensive ways to increase a program's ranking is to fudge the data. This author was not advocating these steps, merely reporting his experiences—and cautions that these suggestions may seem tongue in cheek, but are only a small portion of what deans have described to him over the years. He indicates that "rankings have damaged the business school industry by causing serious misallocation of resources."<sup>61</sup>

Unfortunately, buildings are built, admissions policies set, and services rendered in order to achieve a top twenty-five ranking with oftentimes little consideration actually given to the learning environment of a program.

## THE TEST AND TESTIMONY OF REPUTATION

College ratings and rankings constitute a test of reputation. What have we learned about reputation and how it is formed? The first and most obvious point is that the formation takes time. We might not be surprised, then, that older institutions are better known and represented in many studies of reputation. Indeed, an institution that has been serving this nation for a hundred years or more has earned its way.

Second, size can combine with time to produce a larger number of graduates. When these graduates occupy positions of responsibility, one is also not surprised

to find his or her alma mater well represented in reputational studies. And as the production levels of other American universities grow, we will not be surprised to note some expanding diversity in the institutions appearing in those ratings.

Third, national reputations associate more with publication than with teaching, more with article and book citations than with the knowledge and pleasure of one's graduates. We are not surprised that institutions attracting frequently published research scholars are favored, certainly in graduate reputational studies.

Finally, reputations can be built on some sense of distinction in curriculum or climate. We should also not be surprised that institutions are more easily recognized for quality if they have such a distinction: great books at St. John's, expository grades at Evergreen, work emphasis at Berea, general education competencies at Alverno.

Not appearing in reputational studies is a large range of institutions that are perhaps younger, smaller, less selective, a little more routine in climate and curriculum, but not necessarily of low quality. Those institutions and programs occupying the top positions in reputational studies have no justification to hold in disdain or arrogance those not appearing there. The halls of history are filled with the sounds of bare feet going up the stairs and golden slippers coming down. Graduates of Harvard, Chicago, Yale, Columbia, and Stanford will be working for graduates of Melrose A&M or Siwash College. And researchers from more prestigious schools may be scooped by an obscure scholar working in one of the "academic Siberia" laboratories. All of this gives life to the power of diversity in American higher education.

There is, by the same token, no justification for those not appearing in reputational studies to hold jealousy or prejudice against their colleague institutions for these accomplishments and recognitions. Not only achievement but daring and perseverance are represented in these ratings. With institutions as with individuals, there's little to gain in throwing oneself against the record or reputation of another. To be inspired by the lessons of risk, imagination, achievement, and perseverance demonstrated by those programs and institutions that have earned their way to prominence is a healthy response. To understand that recognition is to be earned by time and distinction is also a healthy response.

With these reflections in mind, we can begin to understand some of the constructive contributions made to quality in American higher education by ratings and rankings, the lifting effect of reputational studies.

### Keeping the Concern for Quality Visible and Active

Certainly reputational studies serve this rather fundamental and simple purpose. Hutchins once said something to the effect that a first-class university is essentially an ongoing and lively argument. Conversations on the nature of quality open the possibility of new action and understanding.

### Reflecting the Power of Innovation

A careful review of the ratings studies, particularly those of undergraduate programs and institutions, illustrates the variety of curricular and program innovations recognized in those studies. These experiments, some small and others large, can furnish both information and inspiration for other campuses. This is a worthy outcome, and reputational studies deserve a salute for that purpose alone.

### Demonstrating the Power of Perseverance

A review of reputational studies helps develop a greater appreciation for the long and distinguished history of some of our national institutions. In a nation that thrives on everything "instant"—from coffee to personal gratification—important lessons are to be learned in the power of perseverance. Northeast Missouri State University has been recognized in the reputational studies and in other places for its daring approach to value-added instruction. What many folks do not recognize, however, is that Northeast has been developing this program since the early 1970s. The university earned this recognition over a long period of time.

### Creating a Competitive Edge

We know both the advantages and dangers of competition in our society. An overemphasis on competition delivers a "dog-eat-dog" mentality in which overly ambitious personalities and institutions climb over the backs of colleagues to fulfill their selfish desires. On the other hand, the absence of the competitive edge can cause both personalities and organizations to lapse into a stupor of mediocrity or explode into arrogance.

These summary reflections on the strengths and weaknesses of reputational studies may leave some readers yearning for a set of one-armed authors, so that equivocation is not encouraged by the phrase "on the other hand." So let us state our conviction clearly: We are not inclined to see reputational studies as a very useful quality assurance tool.

Although scholarly exercises and media studies of ratings can serve to keep the quality argument in ferment and encourage the "culture of evidence" about which we spoke in chapter 2, there continue to be equivocal linkages between quality reputation and quality reality. We find no significant evidence that reputational studies furnish a very meaningful tool for enhancing student choice. Nor do we find that reputational studies make a demonstrable contribution to the most fundamental goal of evaluation and quality assurance—improvement. Finally, we find no obvious value of ratings to the other major quality assurance goal—establishing accountability with major higher education stakeholders.

Reputational studies concentrate on asking questions of those who are responsible for the design and delivery of higher education, but there is a more important respondent. Few Americans who purchase either products or services

in our nation today escape a follow-up telephone call or questionnaire to ascertain customer and client satisfaction. As we shall see in chapter 4, higher education is not a newcomer to the use of client satisfaction indices as an instrument of quality assurance. We pay attention to our clients and customers: our students.

### NOTES

1. McGuire, J., and Others. (1988). "The Efficient Production of Reputation by Prestige Research Universities in the United States." *Journal of Higher Education*, 59, July/August, 356.
2. Zemsky, R., and Massey, W. (1990). "The Rain Cometh." *AGB Reports*, 32(1), 21.
3. Graham, A., and Thompson, N. (2001). "Broken Ranks." *Washington Monthly*, September, 10.
4. News Bureau. (1997). "UI Ranked Second in the Midwest for Quality at a Low Price by Money Magazine." News Release. Available on-line: <http://www.news.uiuc.edu/archives/97.08/14moneymag.num>.
5. *Time*. (1988). September 12, 72-73.
6. Wilson, L. (1964). *The Academic Man*. New York: Octagon Books.
7. Hughes, R. (1925). *A Study of the Graduate Schools of America*. Oxford, OH: Miami University, 3.
8. *Ibid.*
9. *Ibid.*
10. "Report of the Committee on Graduate Instruction." (1934). *Educational Record*, April.
11. *Ibid.*, 194.
12. Keniston, H. (1959). *Graduate Study and Research in the Arts and Sciences at the University of Pennsylvania*. Philadelphia: University of Pennsylvania Press, 115.
13. Cartter, A. (1966). *An Assessment of Quality in Graduate Education*. Washington, DC: American Council on Education, vii.
14. *Ibid.*, 4.
15. *Ibid.*, 3.
16. *Ibid.*
17. Roose, K., and Anderson, C. (1970). *A Rating of Graduate Programs*. Washington, DC: American Council on Education, 2.
18. *Ibid.*, 24.
19. Dolan, W. (1976). *The Ranking Game: The Power of the Academic Elite*. Lincoln: University of Nebraska Press, i.
20. Clark, M., Hartnett, R., and Baird, L. (1976). *Assessing Dimensions of Quality in Doctoral Education: A Technical Report of a National Study in Three Fields*. Princeton, NJ: Educational Testing Service, 1.2.
21. Jones, L., Lindzey, G., and Coggeshall, P. (eds.). (1982). *An Assessment of Research-Doctorate Programs in the United States*. 5 vols. Washington, DC: National Academy Press, v.

22. Ibid., 24.
23. Webster, D. (1984). "America's Highest Ranked Graduate Schools, 1925-1982." *Change*, 15(4), 14.
24. Ibid., 16.
25. Webster, D. S., and Skinner, T. (1996). "Rating Ph.D. Programs: What the NRC Report Says ... and Doesn't Say." *Change*, 28(3), May/June, 26.
26. Ehrenberg, R. G., and Hurst, P. J. (1996). "The 1995 NRC Ratings of Doctoral Programs: A Hedonic Model." *Change*, 28(3), May/June, 46.
27. Webster and Skinner, "Rating Ph.D. Programs," 26.
28. Rogers, E., and Rogers, S. J. (1997). "'High Science' vs. 'Just Selling Magazines': How the NRC and U.S. News Graduate Rankings Compete." *AAHE Bulletin*, 49(9), May, 9.
29. Margulies, R., and Blau, P. (1973). "America's Leading Professional Schools." *Change*, November, 22.
30. "America's Best Graduate Schools." (1990). *U.S. News and World Report*, March 19, 46-78.
31. Ibid., 50.
32. Ibid.
33. Solmon, L., and Astin, A. (1981). "Part One: Departments Without Distinguished Graduate Programs." *Change*, 13(6).
34. Ibid., 27.
35. Solmon, L., and Astin, A. (1981). "Part Two: The Quality of Undergraduate Education: Are Reputational Ratings Needed to Measure Quality?" *Change*, 13(7), 19.
36. Ibid., 18.
37. "Rating the Colleges." (1983). *U.S. News and World Report*, November 28, 41.
38. "America's Best Colleges." (1985). *U.S. News and World Report*, November 25, 46-60.
39. "America's Best Colleges." (1987). *U.S. News and World Report*, October 26, 49-87.
40. Ibid., 70.
41. Ibid.
42. "America's Best Colleges." (1988). *U.S. News and World Report*, October 10, C3-C32.
43. "America's Best Colleges." (1996). *U.S. News and World Report*, September 16, 89-123.
44. "America's Best Colleges." (1999). *U.S. News and World Report*, August 30, 63-105.
45. Ibid.
46. Ibid.
47. "America's Best Colleges." (2000). *U.S. News and World Report*, September 11, 90-133.
48. Ibid., 104.
49. "America's Best Colleges." (2001). *U.S. News and World Report*, September 17, 88-116.

50. Gilbert, J. (1990). "Money's College Value Rankings." *Money Guide*, Fall, 72.
51. "Best College Buys." (2001). *Money*, September.
52. Willard, A. (2001). "Law School Rankings: Through the Education and Employment Looking Glass." Available on-line: <http://www.nalp.org/Schools/rank1.htm>.
53. Webster, D. (1984). "Who Is Jack Gourman and Why Is He Saying All Those Things About My College." *Change*, 13(7), 55.
54. Hutchins, R. (1966). "First Glimpses of a New World." In *What I Have Learned*. New York: Simon and Schuster, 178.
55. Johnson, R. (1978). "Leadership Among American Colleges." *Change*, 10(10), 50.
56. Lawrence, J., and Green, A. (1980). *A Question of Quality: The Higher Education Ratings Game*. AAHE-ERIC/Higher Education Research Report No. 5, Washington, DC: American Association for Higher Education, 10.
57. Caplow, T., and McGee, R. (1958). *The Academic Marketplace*. New York: Basic Books, 14.
58. Gardner, J. (1984). *Excellence*. New York: W. W. Norton, 102.
59. Webster, D. (1981). "Advantages and Disadvantages of Methods of Assessing Quality." *Change*, 13(7), 22.
60. Polocano, A. J. (2001). "Ten Easy Steps to a Top-25 MBA Program." *Selections*, Fall, Graduate Management Admission Council, 40.
61. Ibid.