Boston College

The Graduate School of Education

Department of Educational Research, Measurement and Evaluation

QUANTIFYING QUALITY: TECHNICAL ISSUES SURROUNDING *US NEWS AND WORLD REPORT'S* RANKINGS OF US COLLEGES AND GRADUATE PROGRAMS

Dissertation

By

MARGARET M. CLARKE

Submitted in partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

May 2000

UMI Number: 9961608

UMI®

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

BOSTON COLLEGE

GRADUATE SCHOOL OF EDUCATION

The dissertation of

Margaret Mary Clarke

titled Quantifying Quality: Technical Issues Surrounding US News and World Report's Rankings of US Colleges and Graduate Programs

submitted to the Graduate School of Education in partial fulfillment of the requirements for the degree of:

Doctor of Philosophy

has been read and approved by the Committee:

[signature]

[signature]

[signature]

Date 2/28/00

ABSTRACT

QUANTIFYING QUALITY: TECHNICAL ISSUES SURROUNDING *US NEWS AND WORLD REPORT'S* RANKINGS OF US COLLEGES AND GRADUATE PROGRAMS

By

Margaret M. Clarke

Directed by

George F. Madaus

This study identifies three concerns voiced by critics of *US News and World Report's* rankings of colleges and graduate programs (i.e., the extent to which the rankings represent the relative academic quality of institutions, are comparable across years, and are free from error) and links them to three technical issues (i.e., unidimensionality, comparability, and freedom from error). These technical issues are examined using data from five consecutive editions of *US News'* college and graduate program rankings publications. Results indicate that the indicators used to create the different rankings generally show a unidimensional structure, providing support for the use of a single score to summarize a school's performance, but no information on whether this score actually represents a school's academic quality. Results also indicate that different rankings experienced varying amounts of change in their ranking methodology over the five editions, suggesting that some rankings are more comparable across years than others. This has implications for the interpretation of change in a

school's rank across years. Different rankings also show varying amounts of error around the overall scores for schools, with the education rankings exhibiting the largest amount of error. The amount of error involved is sufficiently large to require considerable caution in the use of overall scores to assign schools to discrete ranks.

ACKNOWLEDGMENTS

I wish to thank my committee members – Drs. Beaton, Madaus, and Shore - for the invaluable advice and support that they extended to me throughout this process. In particular, I wish to thank my committee chair and advisor, George Madaus, for all he has done for me over the last five years. Thanks also to Sol Mumey who did the programming for the freedom from error analyses. Finally, thanks to my friends and family who were always there to listen, talk, or just make me laugh.

Work on this dissertation was made possible by a grant from the Ford Foundation.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter One

Introduction

Background and Statement of the Problem

Before making an expensive purchase, many of today's consumers like to obtain comparative information on the different brands or options available. The consumer guides that contain this information often rank products according to the extent to which they possess various desirable attributes – e.g., value for money, quality, or safety. For the most part, a consumer will choose among products ranked higher, rather than lower on the list, unless performance in a particular area is more important than the overall ranking. Product makers also use these rankings as a way to encourage sales. For example, car manufacturers routinely advertise the fact that their product was ranked first in safety or overall appeal by *Motor Trend, Car and Driver* or similar publications. Thus, both producer and consumer draw on the ranking information contained in these publications.

Rankings are perceived as particularly important when a product is expensive, when many alternative brands or options are available, or when the choice will have long-term repercussions. The process of choosing a college or graduate school fits all of these criteria (Becker, 1993; Johnson and Neal, 1998; Katchadourian and Boli, 1994). Thus, it was probably inevitable that the idea of rankings would be applied to the area of higher education. There are currently several sources of comparative information on higher education institutions, including the popular press, academic journals, and national research organizations. However, students and parents are most likely to be aware of the academic quality rankings produced by magazines such as *Business Week* and *US News and World Report*. The list of publications producing some form of college rankings is growing. Recently, *Newsweek*, *Time*, and *Kiplinger* magazines also jumped on the ratings bandwagon.

The *US News and World Report* (hereafter referred to as *US News*) rankings of colleges and graduate schools are probably the best known of the media rankings. They are also the oldest. *US News* published its first rankings of the academic quality of colleges in 1983. They were immediately popular, if somewhat controversial, and soon became an annual feature. In 1987, *US News'* first academic quality rankings of graduate schools appeared but these were not published on an annual basis until 1990. Sixteen years after their first publication, the rankings (both undergraduate and graduate) are among *US News'* top issues in terms of sales generated (K. Crocker, personal communication, March 19, 1999). In fact, according to some sources, the college rankings publication alone sells about 3.5 million copies on the newsstand, generating revenues of $5.2 million a year (Machung, 1998). These sales figures are testimony to the enormous demand that exists for quantitative, comparable information on the quality of higher education institutions.

This public demand for college and graduate school rankings has made them a topic of interest for researchers and survey groups (e.g., Art and Science Group, 1995; McDonough, Antonio, Walpole and Perez, 1998). It has also made them a focus for much criticism and debate – especially among the institutions that are the subject of the rankings. In addition to questioning the overall concept of ranking higher education institutions, much criticism has focused on the methodology used by *US News* and others to produce the rankings. Gerhard Casper, Chancellor at Stanford, summarizes these methodological concerns in a letter of protest he wrote to the editor of *US News* in 1996:

> Could there not, though, at least be a move toward greater honesty with, and service to, your readers by moving away from the false precision? Could you not do away with rank ordering and overall scores, thus admitting that the method is not nearly that precise and that the difference between #1 and #2 - indeed, between #1 and #10 - may be statistically insignificant? Could you not, instead of tinkering to "perfect" the weightings and formulas, question the basic premise? Could you not admit that quality may not be truly quantifiable, and that some of the data you use are not even truly available (e.g., many high schools do not report whether their graduates are in the top 10% of their class)? Parents are confused and looking for guidance on the best choice for their particular child and the best investment of their hard-earned money. Your demonstrated record gives me hope that you can begin to lead the way away from football-ranking mentality and toward helping to inform, rather than mislead, your readers.[1]

Casper's questions about the "football ranking mentality" employed by *US News* and others go to the heart of the current debate over college and graduate school rankings. If, as Casper states, "the difference between #1 and #2 - indeed, between #1 and #10 - may be statistically insignificant," what are the implications for the way in which the overall scores for schools are used to put them in rank order? In addition, if the weights and formula are constantly being "tinkered" with, how should one then interpret change in a school's rank from year to year? And if "quality may not be truly quantifiable," what

[1] The full text of this letter is available at:
http://www-portfolio.stanford.edu:8050/documents/president/961206gcfallow.html

are the implications in terms of using a combination of indicators and weights to come up with an overall quality score?

Surprisingly, there does not appear to be much research aimed at answering Casper's questions. A 1997 study by Rogers and Rogers addressed the issue of validity in relation to the *US News* rankings of doctoral programs, but most research focuses on the extent to which the rankings are used by students and parents (e.g., Art and Science Group, 1995; McDonough et al., 1998) or their effect on institutions (e.g., Machung, 1998). In fact, what has been termed "the first empirical study" of the *US News* rankings, did not appear until 1999 (Blair, 2000). In this study, researchers examined several years of rankings data and found that a college's place in the rankings was directly related to the number and quality of students who were subsequently admitted and enrolled and the amount of financial aid the college handed out (Monks and Ehrenberg, 1999).

Because of the lack of such empirical studies, the technical issues at the core of Casper's questioning – i.e., the extent to which the rankings produced by *US News* and others represent the academic quality of an institution, are comparable across years, and are free from error – are not being addressed. Empirical studies that address these issues are needed because they can provide concrete information on what rankings can and cannot tell us about the academic quality of an institution. In addition, such studies could provide information that would help address the main question posed by Casper and others – that is, whether the concept of academic quality rankings is itself an appropriate one.

Purpose of the Study

This study examines the technical issues at the core of Casper's concerns – i.e., the extent to which the rankings produced by *US News* represent the academic quality of institutions, are comparable across years, and are free from error. These technical issues are important to address, not only because they underpin the concerns of those critical of the rankings, but also because they are central to the assumptions underpinning *US News'* ranking methodology. These assumptions are as follows:

1. There is a fairly commonly agreed upon set of information that can be used to describe the academic quality of US colleges and graduate schools.
2. This information can be quantified and organized into discrete indicators.
3. These indicators can be related or linked to each other in some fashion (e.g., by categorizing and weighting).
4. The results of this relating or linking can be summed to produce a composite score.
5. This score is a quantitative representation of the qualitative attribute of academic quality.
6. This score can be used to rank order institutions according to the amount of academic quality they possess.

These six assumptions are not specifically stated by *US News* but flow from the steps in their ranking methodology – i.e., *US News* collects information about schools on a set of indicators, weights this information, and adds it together to produce a composite score that is used to rank schools. The final product is called an academic quality ranking. These assumptions are briefly discussed below in terms of the extent to which they relate to the technical issues raised by Casper's concerns and how these issues will be explored in this study. More detailed discussion is reserved for Chapter Three.

Unidimensionality

Assumptions 5 and 6 rest on the idea that a mathematical model – such as that suggested by Assumptions 1 through 4 - can produce a single number that is a representation of the construct or attribute of academic quality. In addition, any difference in the number or score assigned to two schools is assumed to be due to differences in the amount of attribute that they possess – i.e., the main reason that one school receives a score of 86 and another a score of 68 is because the former has more "academic quality" than the latter.

A full evaluation of these assumptions would require a validity study that addressed the extent to which the rankings actually represent the academic quality of an institution. This study will carry out one possible aspect of such an effort by assessing the number of constructs or dimensions that underlie the set of indicators used to produce the overall score. If only one main dimension or construct is identified, the set of indicators is said to be unidimensional. If more than one main dimension or construct is identified, the set of indicators is said to be multidimensional. The number of dimensions that underpin a set of indicators has implications for the way in which these indicators can be combined and interpreted. In terms of the *US News* rankings, if a main dimension were found to underlie the set of indicators used for each ranking, this would provide support for the use of an overall score to summarize the information in the set of indicators. If more than one main dimension were found, it might be better to provide a score for a school on each dimension. Certainly, in this instance, more analyses would be required in order to understand more fully the relationships among the indicators before combining them. It is important to note that while these analyses may tell us the number of dimensions that underlie the overall score for schools, they will not reveal whether the dimension(s) represent(s) academic quality.

Comparability

Drawing on Assumption 1 in particular, if there is a fairly agreed upon set of information that can be used to describe the academic quality of schools, we would expect this information to be employed whenever academic quality rankings are created. In addition, we would then expect change in a school's ranking over time to be attributable to change in its (or other schools') performance on this set of information.

This interpretation of Assumption 1 as requiring comparability across time in ranking formula is stricter than *US News'* interpretation. *US News* acknowledges that they adjust their formula on an annual basis in response to the availability of new sources of information or in an effort to improve the methodology or definition for an existing indicator. The underlying message is that this change is sufficiently small to avoid changing the nature of the overall attribute on which institutions are ranked. Given *US News'* ready acknowledgement of yearly changes in their formula, this study will examine the actual extent of change across years in the information used to construct the *US News* academic quality rankings and the implications of this for the interpretation of change in a school's ranking from year to year.

Freedom from Error

Assumptions 4 and 6 in particular suggest that the data and methods used in the rankings are relatively free from error. In particular, Assumption 6 assumes that the amount of error in the overall score is small enough for it to be used to place schools in a descending order without any overlap. The only overlap that occurs is in the case of schools that receive the same overall score and are thus tied for rank. *US News* has softened this assumption in recent years by rounding overall scores to the nearest whole number – in recognition, they say, of the fact that small differences in scores may be attributable to error (Thompson and Morse, 1998, p.66).

These assumptions about the degree of error in the scores used to produce the rankings can be evaluated in two ways – sources of error and ways of controlling for or measuring error. The most common sources of errors mentioned in relation to the *US News* rankings are computational miscalculations and mistakes in data sent in by institutions (deliberate or otherwise). These sources of error are controlled for or eliminated by *US News* through a process of rechecking calculations and crosschecking data sent in by institutions. This study looks at another source of error that *US News* does not control for – changes in the set of indicators used to compute the overall score – and uses this as a means of estimating the amount of error around the overall score for each school. The amount of error found has implications for how one interprets a school's score – and by extension – its rank. For example, an error estimate of 3 points around a school's overall score of 85 would suggest that the school's score could be as low as 79 and as high as 91 (using a range of approximately ± two errors and assuming a normal distribution of scores) – putting it in the same range as some institutions listed above or below it in the rankings. This in turn could mean that schools that appear to be separated by as many as three or four ranks might in fact have no significant difference in their overall score.

The results of the analyses in each of these three areas - unidimensionality, freedom from error and comparability - will be discussed in terms of the extent to which they support the assumptions underpinning *US News'* ranking methodology and the extent to which they address the concerns raised by Casper and others regarding the appropriateness of academic quality rankings for the evaluation of institutions of higher education.

Significance of the Study

The significance of this study resides in its contribution to the current debate over college and graduate school rankings, its potential application to the ranking of educational institutions in general, and its practical value for the average consumer of rankings information.

Contribution to the Current Debate Over College and Graduate School Rankings

Much of the current debate over college and graduate school rankings is focused on the overall concept of rankings and their appropriateness in the realm of higher education. This approach is evident in the 1998 appeal by the Association of American Law Schools (AALS) to put an end to the *US News* rankings on the grounds that they were "misleading and dangerous," among other things (*US News and World Report*, 1998). This approach is also evident in attempts by Alan Stone, the president of Alma College in Michigan, to drum up interest among college presidents in a national boycott of the *US News* rankings. Stone has criticized *US News'* choice of indicators but his main approach has been an argument against the whole enterprise of ranking schools like "automobiles or toasters."

The reactions of the AALS and Alan Stone exemplify the fact that most criticisms of the *US News* rankings to date have been focused at the conceptual, rather than the technical, level. However, critical analyses at both levels are necessary if the various points of disagreement are to be discussed and resolved in a fully informed manner. This study will provide a first step towards creating articulation between both levels by conducting technical analyses of the assumptions upon which the *US News* rankings are based. The results of these analyses will provide concrete information to inform the conceptual debate.

Potential Application to the Ranking of Educational Institutions in General

At present, almost every state in the US has a statewide accountability system that uses test scores and other indicators to record student, school, and district progress towards specified reform goals. Such systems often end up as rankings of schools and districts – either based on a single indicator such as test scores, or a combination of indicators such as test scores, dropout figures, and graduation rates (Gledhill, 1999).

Many states (e.g., Massachusetts) are currently struggling with whether and how they should combine these various pieces of information in order to make comparisons within educational units (e.g., schools and districts) over time and across educational units on a yearly basis (Cassidy, 1999). Some of the indicators involved in making these comparisons (e.g., test scores and graduation rates) are similar to those used in college and graduate school rankings. Some of the methodological issues involved are also similar – e.g., how to handle possible sources of error in the data; the effects of error on the interpretation of overall scores; and, the extent to which these scores can be used to make comparisons within and across years. While the current study focuses on rankings at the college and graduate school levels, it is hoped that the examination of some of the common technical issues involved will provide useful information for ranking or comparing educational institutions in general.

Practical Value for the Average Consumer of Rankings Information

Rankings are a part of everyday life. The concept of ordering information and objects in some meaningful fashion makes it easier to understand and use information on a variety of products – from cars to computers to cosmetics. Yet how much does the average consumer really know about the way rankings are compiled or the extent to which they are accurate? Most people probably do not stop to think about the indicators chosen to describe a product or the way in which these are combined to come up with an overall ranking of competing products according to some desirable attribute. The choice of these indicators and weights is a value-laden decision and there may be less consensus on the indicators and weights in some areas than in others (Camilli and Firestone, 2000).

Most consumers also do not stop to think about whether it is appropriate to rank certain items, products, or things. For example, rankings can work well in the product domain as they draw on features intrinsic to the product (e.g., materials used and speed) but do not necessarily work well in other domains such as higher education where the criteria used to rank (e.g., student/faculty ratio, alumni giving rate) may not be intrinsic to the entity itself. Consumers need to think about these issues when making choices in any area that does not have readily visible or agreed upon criteria. Thus, while this study analyzes and discusses some technical issues underlying college and graduate school rankings, it also provides some general insights and conclusions that will help consumers to both evaluate and use this type of comparative information in their daily lives.

Limitations of the Study

The limitations associated with this study revolve around the issue of missing data. This issue can be addressed in terms of the systematic omission by *US News* of certain data on a year-to-year basis and the varying amounts of data missing from different rankings at the graduate and undergraduate level.

Systematic Omission of Data

The main limitation associated with this study is that *US News* does not make public all the data it uses to compile the undergraduate or graduate rankings. Therefore, it is difficult to replicate the rankings produced for any particular year. In addition, there is a pattern of the same type of data being held back on a yearly basis so that data from one year cannot be used as estimates for data missing from another year. For example, the *US News* 2000 law school rankings were compiled using 12 indicators of academic quality. Yet, data on only 9 of the 12 were printed in the accompanying rankings guidebook. Data on the other three indicators - *Average Expenditures per Student for Instruction, Library, and Supporting Student Services*; *Average Expenditures per Student for Financial Aid and other Administrative Expenses*; and, *Total Number of Volumes in the Law Library* - were omitted from the guidebook, and have been omitted from all guidebooks for at least the last five editions. *US News* says that these indicators are omitted because they are of less interest to consumers than some of the other rankings information (R. Morse, personal communication, August 1999). For example, test scores have never been left out of the rankings guidebook at either the undergraduate or graduate level as they are perceived to have high consumer value. While levels of consumer interest and space constraints may dictate which indicators are printed in the ranking guidebooks, *US News* also does not make its database available to researchers and has no policy in place for dealing with such requests. Reliance on other sources to replace this missing information was not feasible for this study as much of the missing data are either not available from such sources or else only in a different format. Therefore, it was decided to carry out the analyses using available data and to address the implications of missing data in the discussion of results.

Varying Amounts of Data Missing from the Different Rankings

Missing data also led to constraints in terms of the number of rankings lists that could be examined. For example, there are generally fewer data missing for the law and business school rankings than the rankings of schools of education. This is mainly because more indicators are used to compute the schools of education rankings (14 measures are used for the education rankings, as opposed to 12 for the law and 8 for the business rankings). Since the same space constraints apply to all the rankings, a smaller fraction of the education indicators wind up on the page than for the law and business school rankings. In addition to this problem, *US News* generally does not print data on all schools in a particular field (e.g., business, education or medicine) but tends to provide data on only the top 50 or so. The law schools are an exception to this rule, with data provided for all accredited law schools.

At the undergraduate level, *US News* provides data on all accredited colleges, with more data given for the top-ranked schools and less provided for schools that fall into the lower tiers. With these constraints in mind, this study will examine the business, education, engineering, law, medicine, and primary-care rankings at the graduate level, and the national liberal arts college and national university rankings at the undergraduate level. The main focus will be on the business, education, law, national liberal arts college, and national university rankings since, in general, the most data are available for these rankings.

Brief Overview of the Study

This dissertation is organized as follows. Chapter One has provided an overview of some of the technical issues surrounding the *US News* academic quality rankings and discussed the significance and limitations of the proposed research in addressing these issues. Chapter Two provides a review of relevant literature in the area of academic quality rankings, focusing on two rankings in particular – the National Research Council's ratings of doctoral-research programs and *US News'* rankings of colleges and graduate schools. Chapter Three outlines a methodology for examining the technical issues surrounding the *US News* rankings. Chapter Four presents the data analyses and their results. Chapter Five discusses these results and presents conclusions and recommendations for future research.

Chapter Two

Review of the Literature

The review of the literature on college and graduate school rankings proceeds as follows. Section One discusses the concept of ranking as a form of evaluation - both in general and as a means of evaluating the academic quality of institutions of higher education. Section Two covers the origins of academic quality rankings and describes past and current methods for producing these rankings. Section Three follows the development of one academic quality ranking in particular - the *US News* academic quality rankings of colleges and graduate programs. The *US News* ranking methodology is described and compared to what are considered by many to be the best academic quality rankings to date - the National Research Council's ratings of doctoral-research programs (Jones, Lindzey & Coggeshall, 1982; Goldberger, Maher & Flattau, 1995). Section Four provides an overview of the reactions by various groups to the *US News* rankings and discusses the extent to which the issues of unidimensionality, comparability, and freedom from error are addressed in this conversation. Section Five summarizes the main issues discussed and links them to emerging issues and trends in the field of academic quality rankings.

Ranking as a Form of Evaluation

The evaluation process normally involves identification of relevant standards of merit, value, or worth; some investigation of the performance of the evaluand on these standards; and some integration or synthesis of the results to achieve an overall evaluation or set of related evaluations (Joint Committee on Standards for Educational Evaluation, 1994). For example, the letter grade that sums up a semester's work by a student describes the quality of that work.

According to Scriven (1991), ranking involves "placing individuals in an order, usually of merit, on the basis of their relative performance on (typically) a test or measurement or observation" (p. 299). Thus, ranking is a type of evaluation in that it is a process of determining the relative merit, value, or worth of something, or the product of that process.

Ranking as evaluation is very common within the world of product evaluation. Perhaps the best-known purveyor of this type of information is *Consumer Reports* which has been rating products since 1936. Product evaluations generally begin with facts about the performance of various products and draw conclusions about their relative or absolute merit. As Scriven (1991) states:

> The performance of the entity to be evaluated is analyzed into a set of dimensions that are independent and exhaustive and, preferably, familiar from other contexts or easily grasped...The dimensions may include some that are descriptive...others that are evaluative, and many that are implicitly ... evaluative...(p. 124).

After performance on each dimension has been rated, the set of ratings is usually combined into an overall rating. Weight-and-sum is the dominant model for this combinatorial step. Usually, the dimensions of merit are weighted for their relative importance and then points are awarded for a product's performance on each of these dimensions. The sums of the weights and performance scores are totaled for each product, the best product being the one with the highest total.

Critics have pointed out several problems with this weight-and-sum model (Evaluation News, 1981). One problem is that no set of weights can cover the situation in which a minimum performance on some of the dimensions is essential. Another problem is that the choice of weights is a value judgement and thus could vary depending on who makes the decision (Camilli and Firestone, 2000). Weighting is also affected by the number of criteria. Depending on the interaction between the number of criteria and their weights, one dimension may dominate all others or several trivial dimensions may swamp more crucial ones.

There is also no consensus on whether the weight-and-sum model translates well into other domains such as personnel or program evaluation. For example, critics charge that the movement from facts to evaluative conclusions works in a product evaluation context because it relies on shared values among its readers, and this will not transfer to, say, the evaluation of higher education institutions. However, Scriven (1991) makes an argument for the use of this type of evaluation in other arenas as follows:

> Our language implicitly defines ideal types in the product field, as it often does in the psychological and sociological field, and we use them...as the norms against which we rate actual products. The ideal types themselves are based on functional and definitional analysis, not on popularity polls. The same model we use in product evaluation applies - with minor modifications - to candidates via job descriptions...and to social programs in a similar way (p. 218).

While Scriven's argument highlights the apparent ease with which the weight-and-sum product evaluation model can be applied to other arenas, the variety of ways in which institutions of higher education are currently evaluated shows no such consensus on this issue.

Evaluating Educational Institutions

According to *The Program Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 1994), education and training programs are primarily evaluated "...in order to determine their quality and gain direction for improving them..." (p. 1). Thus, it is not surprising that one of the main criteria used for evaluating institutions of higher education is academic quality.

The higher education community evaluates academic quality through its own accrediting process. In the US, a college or university must meet a set of standards and goals laid out by one of several regional or national accrediting agencies in order to be recognized as fully accredited. Institutions are compared to a set standard and not to each other. An institution either meets or does not meet the standard, and thus can be placed in one of two, possibly three, groups - accredited, not accredited, on probation/in review. Since the accreditation process is primarily a source of information for schools, students, parents and other consumers generally go elsewhere to obtain information on institutions. There are several sources available - ranging from "non-evaluative" to "evaluative" in nature. At the non-evaluative end, every school produces its own set of written materials describing its campus and programs to prospective students. These materials usually combine a mix of statistics, narrative and visuals and are designed to describe and market the individual school, not compare it objectively to others or to an independent, external standard of quality. College and university websites and videos repeat or expand on this information.

Slightly more evaluative in nature are commercial college and graduate school guidebooks. These generally list all accredited colleges and universities in the US and include some factual information on each - e.g., acceptance rates, average test scores of the previous year's incoming freshmen, and graduation rates. Some guides, such as Barron's *Profiles of American Colleges* and Peterson's *4 Year Colleges*, place schools into four or five levels of selectivity based on a combination of average test scores, class rank and percent accepted for the previous year's freshmen class. These selectivity levels provide some degree of stratification but avoid ordering individual institutions.

Even more evaluative in nature are the college and graduate school rankings produced by popular magazines such as *Business Week*, *Kiplinger's*, and *US News*. Using variations on the weight-and-sum model of product evaluation, criteria ranging from test scores to graduation rates to faculty resources are combined to produce a rank-ordering of institutions - usually according to their academic quality or value for money. This approach allows consumers to compare directly one institution against another; it does not allow them to compare schools against an independent, external standard of academic quality. Other sources of rankings information include websites that provide links to various types of college and graduate school rankings (e.g., Boston College has a website devoted to the topic of rankings with links to 22 sites covering graduate, undergraduate, law, business, and international rankings).[1] Also available in some high school and college libraries are compilations of educational rankings known as ranking annuals that list rankings in hundreds of areas, covering many different types of institutions in several countries (e.g., Hattendorf, 1993).

Ranking as a Way of Evaluating Educational Institutions

While the concept of product rankings is relatively acceptable to most, there is no consensus on the ranking of higher education institutions. For example, some have claimed that "there is a madness in trying to rank institutions...There are no data I know that can judge the culture of an institution" (Webster, 1992, p.19). On the other hand, others have acknowledged that "the public wants anything that will rate and compare colleges" (Carmody, 1987, p.19). The tension here is between a conceptual argument against rankings and a perceived public need for them.

Hattendorf (1993) explains that compiling educational rankings is a complex process because they attempt to measure an intangible known as "quality." While product rankings are compiled on objective measures intrinsic to the product such as (in the case of a car) acceleration, braking, and fuel economy, academic quality rankings are often based on more subjective, less easily interpretable, measures such as academic reputation, admissions selectivity, application rates, distinguished alumni, student achievement in later life, and test scores. These serve as indirect indicators of academic quality - they are not measuring devices and are not intrinsic to the "product" being

---

[1] The address for this website is: http://www.bc.edu/bc_org/avp/ulib/ref/guides/edu/educrankings.html

evaluated. Thus, it is not as easy to reach consensus on which measures and weights to use when assessing academic quality as it is in the case of assessing car performance.

Nonetheless, many believe academic quality rankings to be useful as they fill a public need for comparable information on institutions of higher education that accrediting agencies, commercial guidebooks, and colleges do not meet (Crenshaw, 1999; Sanoff, 1998; Webster, 1986a; Webster, 1992; Wright, 1990-91). The media have picked up on this public need. When asked whether schools should be ranked (and whether this should be the job of the media), a former *US News* editor noted that, absent an agreement among the higher education community on what a successful outcome is, the media rankings at least provided some sort of accountability and measures. He concluded that if higher education were more willing to perform this function, there might not be a need for media rankings (A. Sanoff, personal communication, August, 1999). What is interesting is the way the media have now gained control over the way in which higher education institutions are evaluated and the type of information made available to consumers. As Wright (1990-91) puts it:

> We in the profession are allowing journalists to measure the quality of institutions and education [in the form of rankings], to act as authorities in a field that is much more complex than they understand. And this, we must admit, is largely our own fault, because we do not ourselves have consistent measures of quality (p. 16).

As a form of protest against the media rankings, some institutions are now providing the information themselves to consumers. For example, Stanford and a few other universities have placed the statistical information they send to *US News* on their own websites and encouraged other schools to do the same (Richtel, 1998; Thompson, 1996).[2] However, the general lack of consistent and comparable information on colleges and graduate schools, and the public need for such information, has, at least in part, legitimated the use of rankings in the eyes of many (Gilbert, 1992). As Webster (1986a) states:

> Nonetheless, just as democracy, according to Winston Churchill, is the worst form of government except for all the others, so quality rankings are the worst device for comparing the quality of American colleges and universities, except for all the others (p. 6).

The conceptual debate over whether higher education institutions can or should be ranked has shifted focus to the issue of how to rank. There is no consensus in this area either (Hattendorf, 1993). The number of ways to rank schools is as varied as the types of information used to rank them. The next section looks at the origins of higher education academic quality rankings, the various methods used to create these rankings over time, and the perceived advantages and disadvantages of each.

The Origins and Development of Academic Quality Rankings

The origins of academic quality rankings of colleges and graduate schools can be found in the heredity versus environment debate that swept Europe in the nineteenth century (Webster, 1981, 1986a). As part of the discussion over whether geniuses were products of heredity or environment, the ecological origins of eminent men and women were assessed: where they grew up, where they attended school, and whether and where

---

[2] The Stanford website is www.stanford.edu/home/statistics

they attended university. In his 1900 book, *Where We Get Our Best Men*, Alick Maclean provides a rank order of the universities attended by 3,968 eminent men. Of these, 1,230 had attended universities in England, Scotland, or Ireland, as follows:

- Oxford (424)
- Cambridge (381)
- Edinburgh (132)
- London (97)
- Dublin (73)
- Glasgow (52)
- Aberdeen (42)
- St. Andrews' (18)
- Belfast College, University of Ireland (6)
- Durham (3)
- Owens College (2)

This rank ordering of universities that had "produced" distinguished people came very close to being an academic quality ranking. According to Webster (1986a), when academic quality rankings were "invented" in the US, shortly after the turn of the century, they were quite similar to Maclean's in that they were based on the undergraduate and doctoral origins and, especially, on the current academic affiliations of eminent American scientists. These American rankings will be discussed in more detail below.

It is interesting to note that, despite their promising start in Europe, higher education rankings did not grow to become the phenomenon there that they are today in the US. Apparently, the US provided a more conducive environment for these rankings due to the fact that it had a larger number of relatively new universities, without established reputations, that were all competing for the same students (Webster, 1986a). In addition, these American universities varied much more in type, curricular offerings, and quality than their European counterparts. Rankings thus became a useful means of recording the outcome of competition of US universities for students and financial support.

A Definition of Academic Quality Rankings

According to Webster (1986a), an academic quality ranking is one that meets the following criteria:

- It must be arranged according to some criterion or set of criteria which the compiler(s) of the list believed measured or reflected academic quality.
- It must be a list of the best colleges, universities or departments in a field of study, in numerical order according to their supposed quality, with each school or department having its own individual rank, not just lumped together with other schools into a handful of quality classes, groups or levels (p. 5).

Almost all current quality rankings meet the first criterion, but not all meet the second, especially at the elementary and high school levels. For example, while the Academic Performance Index (API) that underpins the state accountability system for Californian elementary and high schools is called a ranking, it groups schools into performance bands based on their score on the API and does not assign individual ranks (Gledhill, 1999). At the college level, most rankings meet both of the above criteria, suggesting a general consensus on the definition of an academic quality ranking, but not necessarily consensus on how to measure academic quality. In fact, a brief review of the different methods used to produce college and graduate school rankings indicates that how you measure academic quality very much depends on how you define it. For example, three common definitions used when assessing quality in higher education are *institutional reputation* (measured through surveys), *resources* (measured through collection of data on library resources, expenditure per student etc.), and *value-added* (measured through comparing expected with achieved outcomes in areas such as retention and student graduation) (Braxton and Nordvall, 1996). Most of the media rankings of colleges and graduate programs incorporate several of these definitions (and associated measures) into their methodology.

Past and Current Methods for Producing Academic Quality Rankings in the US

Attempts to rank US institutions of higher education date to the late nineteenth and early twentieth century (Hattendorf, 1996). Early attempts by organizations such as the United States Bureau of Education (e.g., Babcock, 1911; United States Bureau of Education, 1917) and the American Association's Council on Medical Education (e.g., Flexner, 1910) involved classifying institutions or particular fields of study like medicine into two or more strata according to their presumed academic quality. Other attempts by accrediting bodies for professional fields (e.g., Gies, 1926) were more focused on enforcing minimal standards than with distinguishing among varying degrees of excellence, and generally produced undifferentiated lists of approved schools.

The man who actually "invented" academic quality rankings in the US was the psychologist, James McKeen Cattell. After earning his Ph.D. in 1886, he worked at Francis Galton's anthropometric laboratory in London. Galton's statistical methods and his interest in individual differences influenced Cattell. Upon returning to the US, Cattell focused on his interest in eminent people, in particular scientists. He produced his first reference work, *American Men of Science: A Biographical Dictionary*, in 1906. This work listed more than 4,000 scientists in North America, with an asterisk or "star" beside about 1,000 of the most eminent.

In 1910 Cattell decided that his list of starred scientists needed to be updated. He asked the ten scientists who had been ranked at the top of each of the twelve branches of science in the 1906 edition to make new lists of the leading ten people in their discipline, in order of their scientific eminence. Cattell averaged the rankings of those who responded, forming a rank order of the leading scientists in order of their scientific merit for the same twelve scientific fields. Cattell then devised a complex weighting system to derive from the number of starred scientists at leading universities and other organizations employing large numbers of starred scientists the first academic quality ranking (see Cattell, 1910b, 1910c). Here, in his own words, is how he did it:

> In order... to sum up in one figure the strength of a university or department, weights have been assigned to the men [that is, the 1000 starred scientists] on this basis - a man in the lower four hundred being the unit, those in the other hundreds were assigned ratings as follows: VII and VI = 1.2; V = 1.4; IV = 1.6; III = 1.9; II = 2.2; and I = 3. The first hundred were subdivided, the lower fifty being assigned 2.5, and the upper twenty-fives, respectively 3 and 4 (Webster, 1986a, p.468)

This, then, was the first academic quality ranking of US institutions ever published. It was a rank order, not just a classification of schools into a few groups. In addition, it was based on a criterion - how many eminent scientists were connected with each institution -

that Cattell thought reflected academic quality. His rankings, which he entitled "the Scientific Strength of the Leading Institutions," are shown in Table 1.

Table 1
The Scientific Strength of the Twenty Leading Institutions in 1910 According to Cattell[a]

Name of Institution	Weighted Number	Gain or Loss Over Score Received in 1903[b]
Harvard	146.0	+16.3
Chicago	94.6	+18.0
Columbia	79.3	-13.3
Johns Hopkins	63.4	+ 4.2
Yale	61.7	+12.2
Cornell	57.6	+ 4.6
Wisconsin	49.0	+22.3
Geological Survey	43.8	-12.2
Department of Agriculture	40.9	- 4.9
Massachusetts Institute of Technology	37.7	+ 9.5
Michigan	37.1	- 3.5
California	32.4	- 5.0
Carnegie Institution	30.9	+19.4
Stanford	30.0	+ 4.8
Princeton	28.6	+ 7.5
Smithsonian Institution	26.0	- 7.3
Illinois	25.0	+16.7
Pennsylvania	24.4	- 4.5
Bureau of Standards	18.9	+ 0.1
Clark	16.0	+ 2.0

[a]Source: Poffenberger, A. T. (1947). James McKeen Cattell: Man of Science, 1860-1944. Lancaster, PA: Science Press.
[b]Although Cattell's first reference work was published in 1906, the data were gathered in 1903.

Cattell's rankings included universities such as Harvard, Yale, and Princeton, but also non-university organizations such as the Smithsonian Institution and the Bureau of Standards. While all 20 institutions listed in Table 1 could be called prestigious, the weighted numbers shown in column two indicate a very wide range of scores, with some schools separated from their neighbors in the rankings more than others. For example, while there is a difference of 130 points between first-ranked Harvard and twentieth-ranked Clark, most of this range (66.7 points) is concentrated in the difference between top-ranked Harvard and third-ranked Columbia. This illustrates how rankings can be deceptive in that the difference in scores between first- and third-ranked (in this case, 66.7) can vary considerably from the difference in scores between third and fifth (in this case, 17.6), even though there is only a two-rank difference in each case. The information in the third column of Table 1 tells us how an institution's weighted number has changed from 1903. While Harvard is ranked first both years, several other schools' rankings changed during this period.

For decades after Cattell's 1910 ranking, dozens of studies were done ranking colleges and universities according to how many important and accomplished people had studied at them. Whereas Cattell had produced the first rank ordering of colleges and universities by examining his 1000 starred scientists' current affiliations, most of the researchers who followed him produced rankings by showing the number of eminent people in various fields who had attended these colleges and universities as students. This approach was more popular because all colleges and universities could be rated by the graduates they "produced," while only a few could be rated by the research achievements of their faculty members since only a small percentage of the colleges and universities had a research faculty. However, by the mid-1960s, the method of ranking colleges and universities by the number and proportion of eminent alumni they "produced" had declined as several studies (e.g., Astin, 1961, 1962) had shown that the number of eminent alumni a college produced was largely dependent on the intellectual ability of the students it admitted, reputational rankings had became more popular (e.g., Blau and Margulies, 1974-75; Cartter, 1966; Roose and Andersen, 1970), and new kinds of reference works (e.g., the *Science Citation Index* in 1961 and the *Social Sciences Citation Index* in 1966) had appeared that inspired new methods for producing academic quality rankings.

By the 1980s, there were three main types of academic quality rankings (Webster, 1986a): rankings based on *faculty accomplishments*; rankings based on *student achievements*; and, rankings based on *institutional academic resources*. These were produced using the following six methods (either individually or in combination): ratings of faculty or program reputation; counts of faculty awards, honors, and prizes; counts of citations in citation indexes; students' achievements in later life; students' scores on standardized tests; and, institutional academic resources. Arguments appear throughout the literature for and against each of these ranking methodologies (Hattendorf, 1993). The author has summarized some advantages and disadvantages associated with each in Table 2.

Table 2
Main Types of Academic Quality Rankings

Type	Method	Advantages and Disadvantages
Based on Faculty Accomplishments	Reputational surveys	**Advantages** They produce results with face validity, i.e., results that most nearly match what the educated public considers the hierarchy of colleges and universities to be **Disadvantages** The overall reputation of an institution may influence raters' assessment of the particular department they are being asked to rank
	Counts of faculty awards, honors, and prizes	**Advantage** They are useful for ranking the best or better institutions **Disadvantage** They may be years behind or ahead of reality
	Counts of faculty citations in citation indexes	**Advantage** Useful in assessing the influence and importance of faculty members' publications, and not just their sheer volume **Disadvantage** The citation indexes on which the rankings are based do not distinguish between "good," "neutral," or "bad" citations
Based on Student Achievements	Distinguished alumni and graduates' achievements after graduation	**Advantage** While only a small percentage of colleges and universities have faculties that produce much research, almost all of them attempt to prepare their students for rewarding careers in later life **Disadvantage** Usually years, if not decades, behind reality
	Incoming students' scores on standardized tests	**Advantage** The data are easy to obtain and are a measure on which most institutions can be ranked **Disadvantage** Based on the academic abilities of students before they enter college and thus fail to consider anything that these institutions do to educate their students once they enroll
Based on Institutional Academic Resources	Compilation of measures of institutional resources, including educational expenditure per student, faculty-student ratios, and library resources	**Advantage** The data are easy to obtain and are a measure on which all institutions can be compared **Disadvantage** Offers little or no information about how often and how beneficially students use these resources

As Table 2 illustrates, the advantages of rankings based on faculty accomplishments are related to their usefulness in distinguishing among the most prestigious colleges and universities. A common disadvantage is that these rankings are not good at distinguishing among the less prestigious schools. In contrast, the rankings based on student achievements and institutional academic resources have the advantage that they can be used to compare all schools, not just some. However, while the data used appears objective (e.g., test scores and expenditure per student), it does not necessarily have the same meaning across different types of schools.

Preferred Ranking Methodologies at the Graduate and Undergraduate Levels

Focusing on undergraduate rankings published since 1950, the most common methods for ranking undergraduate colleges have been by the number or proportion of their graduates who later achieve eminence and by the scores their entering freshmen achieve on standardized tests. Few reputational rankings have been done at the undergraduate level. Of the five that Webster (1986b) could find, two of them were the reputational rankings that *US News* produced in 1983 and 1985 - the early years of its rankings publications.[3] Multi-dimensional rankings now constitute a large proportion of the media ratings of undergraduate colleges and programs (e.g., *Kiplinger's* and *US News and World Report*). Generally, these rankings are based on some combination of

[3] For example, in 1983, *US News* asked presidents of four-year colleges to name the highest-quality undergraduate schools in each of five categories: national universities; national liberal arts colleges, regional liberal arts colleges, comprehensive universities, and smaller comprehensive universities. It then divided the last three categories by geographic area and listed the highest-ranked schools in each category.

the three types of information identified in Table 2 - i.e., faculty accomplishments, student achievements, and institutional academic resources.

In the ranking of graduate-level professional schools and Ph.D.-granting departments, faculty characteristics - especially scholarly achievement as evidenced by publications, citations, awards, honors and research reputation - have tended to count far more than any other criterion (e.g., Bodenhorn, 1997; Conrad and Webster, 1986; Doctoral Faculty Decade Publication Project, 1999; West, 1978). In particular, reputational rankings have been a well-known and influential form of evaluating Ph.D.-granting departments (Goldberger, Maher, and Flattau, 1995; Jensen and Webster, 1981; Jones, Lindzey and Coggeshall, 1982; Webster, 1986a). Multi-dimensional rankings have seldom been made of either graduate-level professional schools or Ph.D.-granting departments. However, this situation changed in the 1980s with the publication of *US News'* rankings of graduate and professional programs and the National Research Council's ratings of doctoral-research programs.

While there are several types of academic quality rankings and methodologies, the issues of unidimensionality, comparability, and freedom from error apply to most of them. The next section addresses the extent to which these issues have been addressed by those creating or interpreting these rankings.

Academic Quality Rankings and the Issues of Unidimensionality, Comparability, and Freedom from Error

Unidimensionality. Since Cattell produced his first ranking of the scientific strength of the leading US institutions, most rankings have been produced using a single measure or indicator of academic quality (e.g., rankings based on reputation or citation analysis data alone). To many observers this approach might seem lacking since one indicator is unlikely to capture the academic quality of an institution. Multidimensional rankings - those that combine various pieces of information on an institution - would seem to provide an answer. However, one is then left with the issue of how much and what type of information to include, and how to combine and weight it to come up with an overall ranking score. Some multidimensional rankings avoid this issue by stopping short of the final combinatorial step (e.g., the multidimensional ratings produced by the National Research Council which present data on institutional performance on several different dimensions but do not combine this information to come up with an overall score). In contrast, the multidimensional rankings produced by the media tend to use expert opinion and their own value judgements to come up with a set of weights, which they then use to combine the various pieces of information into a composite score. This composite is used to represent the academic quality of an institution.

Different interpretation issues are raised by each of these approaches. In the case of rankings produced using a single indicator, there is the danger of an overly narrow or inadequate representation of the construct of academic quality. In the case of rankings that present several pieces of information but stop short of combining them, there is the danger that the reader will be overwhelmed by too much information and too little evaluation of that information. In the case of rankings that combine indicators to produce an overall score, there is the danger that the relationships among indicators may not warrant such summarization and that scores on smaller clusters of indicators might present a more meaningful picture of a school's performance. In all three instances, there is no guarantee that the information presented or the rankings made based on this information actually reflect institutions' relative academic quality.

Comparability. The fact that most ranking studies have focused on one measure or indicator of academic quality has made it easier for researchers to track a college's change in rank (but not necessarily academic quality) across studies over time. For example, an analysis of six reputational rankings of Ph.D.-granting English departments that have been published since 1925 found great stability over time in the highest-ranked departments with few departments entering or leaving the group (Webster, 1990).

Even if academic quality ranking studies for a certain type of program or discipline have tended to rely predominantly on one type of ranking or another, comparability problems can still exist. This is because these studies tend to be carried out by different people, in different organizations, with differing methodologies. There are some exceptions to this tendency. For example, when Cattell produced his 1910 rankings of the relative scientific strength of leading US insitutions, he was able to show which had increased or decreased in strength since he last collected data in 1903 as his methodology was quite similar in both instances (see Table 1). More recently, due to maintaining consistency between the samples and measures used by their 1982 and 1995 studies, the 1995 National Research Council report on doctoral-research programs was also able to present change measures for several aspects of program quality.

Freedom from error. Most attempts at ranking institutions have not discussed the issue of error at any length, if at all. For example, while reputational rankings are often based on a survey of some or all of the researchers or faculty in a particular area, adjustments are not always made for error introduced by sampling or low response rates. Other possible sources of error in the various ranking methodologies include definitional problems with indicators that may corrupt their meaning and mistakes in data collection, entry, or analysis.

The most important implication of these various sources of error is the effect they may have on the overall score used to rank institutions. Few studies take this into account when presenting and interpreting their findings. As will be discussed later, one exception to this is the 1995 ratings of doctoral-research programs produced by the National Research Council which presented confidence intervals around the ratings of the scholarly quality of program faculty.

The next section discusses two of the best known academic quality rankings - the graduate and undergraduate rankings produced by *US News* and the ratings of doctoral-research programs produced by the National Research Council - and discusses the extent to which they identify and address the issues of unidimensionality, comparability, and freedom from error.

Rankings Produced by *US News and World Report* and the National Research Council

***US News'* Rankings of Colleges and Graduate Schools**

US News' rankings of colleges and graduate schools have grown in complexity and influence since the magazine began them in 1983 (Geraghty, 1997; Shea, 1995). At the start, *US News* ranked colleges every two years; since 1987 it has ranked them every year. The first three college rankings were based entirely on institutions' academic reputations, as perceived by college presidents. Then, partly because many college administrators suggested that objective data would yield more valid rankings, in 1988 the magazine began to base 75% of a school's ranking on various objective indicators and only 25% on reputation. That year it also began to base its reputational data on the opinions of not only college presidents, but also deans and directors of admission. In 1987, *US News* produced the first graduate-level rankings in four fields - business, engineering, law, and medicine. In 1994, it added schools of education. Since 1988, in addition to the magazine rankings, it has published guides that contain these rankings and other material on choosing a college or graduate school. Since 1987, *US News* has changed its rankings formula every year.

Current *US News* College Ranking Methodology

The current method that *US News* uses to produce undergraduate academic quality rankings has three basic steps.[4] First, the approximately 1,400 accredited colleges in the US are placed into one of 11 categories based on mission and region.[5] Colleges within each category are ranked separately. Second, *US News* collects data from each school on up to 16 separate indicators of what it believes reflects academic quality (Table 3). These indicators fall into seven broad categories: Academic Reputation; Retention; Faculty Resources; Student Selectivity; Financial Resources; Alumni Giving; and, (for national universities and national liberal arts colleges only) Graduation Rate Performance. As Table 3 indicates, each indicator is assigned a weight in the ranking formula that reflects the judgement of *US News* about which measures of quality matter most. Indicators within categories are standardized and then combined to produce a ranking for each school on this category - e.g., each school has an Academic Reputation rank, a Student Selectivity rank, and so on. Column 4 of Table 3 shows the weight that each indicator (shown in column 3 of Table 3) receives within its category. For example, a school's acceptance rate is 15 percent of its Student Selectivity category score or rank, while SAT/ACT scores contribute 40 percent to a school's rank in this area. Finally, these categories are weighted and combined and colleges are ranked based on the total

---

[4] For more information see http://www.usnews.com/usnews/edu/college/corank.htm

[5] *US News* uses a modification of the classification system developed by the Carnegie Foundation for the Advancement of Teaching in order to classify colleges and universities The Carnegie system is a generally accepted classification system for higher education. The magazine collapses eight of the Carnegie categories (Research Universities I, Research Universities II, Doctoral Universities I, Doctoral Universities II, Master's/Comprehensive Universities and Colleges I, Master's/Comprehensive Universities and Colleges II, Baccalaureate/Liberal Arts Colleges I, and Baccalaureate/Liberal Arts Colleges II) into four: National Universities, National Liberal Arts Colleges, Regional Universities and Regional Liberal Arts Colleges.

score that they receive. Column 2 of Table 3 shows the weight that each category of indicators receives in the final ranking. For example, a school's score or rank in the area of Academic Reputation is 25 percent of its overall rank, while its score or rank in the area of Financial Resources contributes just 10 percent to the final outcome. *US News* publishes the individual ranks of only the top schools; the remainder is grouped into tiers.

Table 3
***US News* Indicators and Weights for the 2000 College Rankings**[a]

Ranking Category	Category Weight	Indicator	Indicator Weight
Academic Reputation	25%	• Academic Reputation Survey	• 100%
Student Selectivity	15%	• Acceptance Rate • Yield • High School Standing – Top 10% • SAT/ACT Scores	• 15% • 10% • 35% • 40%
Faculty Resources	20%	• Faculty Compensation • Faculty With Top Terminal Degree • Percent Full-time Faculty • Student/Faculty Ratio • Class Size, 1-19 Students • Class Size, 50+ Students	• 35% • 15% • 5% • 5% • 30% • 10%
Retention Rate	20%	• Average Graduation Rate • Average Freshmen Retention Rate	• 80% • 20%
Financial Resources	10%	• Educational Expenditures Per Student	• 100%
Alumni Giving	5%	• Alumni Giving Rate	• 100%
Graduation Rate Performance	5%	• Graduation Rate Performance	• 100%

[a]These indicators and weights are for the national liberal arts and national university rankings only.

As Table 3 illustrates, all three types of rankings described earlier - i.e., rankings based on faculty accomplishments, student achievements, and institutional academic resources (see Table 2) - are combined to produce the final academic quality ranking. Appendix A provides more detail on the indicators that comprise each category and the data collection strategies involved.

Current *US News* Graduate Ranking Methodology

A similar methodology is employed for the graduate school rankings. *US News* collects data from each program on indicators of what it believes reflect academic quality. Each indicator is assigned a weight based on *US News'* judgment about which measures matter most. Schools are ranked within indicator categories (e.g., Placement Success and Academic Reputation) before these are combined to produce an overall rank.

The five major disciplines examined yearly are business, education, engineering, law, and medicine. Master's and doctoral programs in areas such as the arts, sciences, social sciences, library science, and various health fields are ranked only by reputation and are generally evaluated every third year. The specific indicators and weights used for rankings within each of the five major disciplines are outlined in Tables 4 through 9.

As Tables 4 through 9 indicate, the graduate school rankings use fewer indicators than the college rankings and reputation counts for much more of the overall score (40% versus 20% for the undergraduate rankings) since two sets of reputation ratings are employed.[6] Appendix B provides more detail on the indicators used for each ranking and an overview of the general methodology employed.

---

[6] In one set, deans and faculty members are asked to rate the academic quality of schools/programs with which they are familiar on a scale of 1 ("marginal") to 5 ("distinguished"). The second reputational ranking is based on surveys of people outside academia who are likely to hire new graduates or come in contact

Table 4
***US News* Indicators and Weights for the 2000 Business Rankings**

Ranking Category	Category Weight	Indicator	Indicator Weight
Reputation	40%	• Academic Survey • Non-academic Survey	• 60% • 40%
Placement Success	35%	• Median Compensation • Employment at Graduation and Three Months Later	• 40% • 20% and 40%
Student Selectivity	25%	• Graduate Management Admission Test Scores • Undergraduate Grade Point Average • Proportion of Applicants Accepted	• 65% • 30% • 5%

Table 5
***US News* Indicators and Weights for the 2000 Education Rankings**

Ranking Category	Category Weight	Indicator	Indicator Weight
Reputation	40%	• Academic Survey • Non-academic Survey	• 60% • 40%
Student Selectivity	20%	• Average Verbal, Analytic and Quantitative GREs • Proportion of Applicants Accepted	• 30% each • 10%
Faculty Resources	20%	• Ratio of Full-time Doctoral and Master's Degree Candidates to Full-time Faculty • Percent of Faculty Given Awards • Number of Doctoral and Master's Degrees Granted in 1998 • Proportion of Graduate Students Who Are Doctoral Candidates	• 25% and 20% • 20% • 15% and10% • 10%
Research Activity	20%	• Total Research Expenditures • Research Expenditures Per Faculty Member	• 75% • 25%

with them in the workplace. Respondents are asked either to use the same rating system as for deans and faculty or to select the top 20 to 25 programs in their area.

Table 6
***US News* Indicators and Weights for the 2000 Engineering Rankings**

Ranking Category	Category Weight	Indicator	Indicator Weight
Reputation	40%	• Academic Survey • Non-academic Survey	• 60% • 40%
Student Selectivity	10%	• Average Quantitative and Analytic GREs • Proportion of Applicants Accepted	• 45% each • 10%
Faculty Resources	25%	• Ratio of Full-time Doctoral and Master's Degree Candidates to Full-time Faculty • Proportion of Faculty Who Are Members of NAE • Number of Ph.D Degrees Granted in 1997-98 • Proportion of Faculty Holding Doctoral Degrees	• 25% and 10% • 25% • 20% • 20%
Research Activity	25%	• Total Research Expenditures • Research Expenditures Per Faculty Member	• 60% • 40%

Table 7
***US News* Indicators and Weights for the 2000 Law Rankings**

Ranking Category	Category Weight	Indicator	Indicator Weight
Reputation	40%	• Academic Survey • Non-academic Survey	• 62.5% • 37.5%
Student Selectivity	25%	• Median LSAT Scores • Median Undergraduate GPA • Proportion of Applicants Accepted	• 50% • 40% • 10%
Placement Success	20%	• Employment Rates at Graduation and Nine Months Later • Bar Passage Rate	• 30% and 60% • 10%
Faculty Resources	15%	• Average Expenditures Per Student For Instruction etc. • Student to Teacher Ratio • Average Expenditures Per Student For Financial Aid etc. • Total Number of Volumes in Law Library	• 65% • 20% • 10% • 5%

Table 8
***US News* Indicators and Weights for the 2000 Medicine Rankings**

Ranking Category	Category Weight	Indicator	Indicator Weight
Reputation	40%	• Academic Survey • Non-academic Survey	• 50% • 50%
Student Selectivity	20%	• Average MCAT Scores • Undergraduate Grade Point Average • Proportion of Applicants Accepted	• 65% • 30% • 5%
Faculty Resources	10%	• Ratio of 1998 Full-time Science and Clinical Faculty to Full-time Students	• 100%
Research Activity	30%	• Total Dollar Amount of National Institutes of Health Research Grants Awarded to the Medical School and its Affiliated Hospitals, Averaged for 1997 and 1998	• 100%

Table 9
***US News* Indicators and Weights for the 2000 Primary-care Rankings**

Ranking Category	Category Weight	Indicator	Indicator Weight
Reputation	40%	• Academic Survey • Non-academic Survey	• 62.5% • 37.5%
Student Selectivity	15%	• Average MCAT Scores • Undergraduate Grade Point Average • Proportion of Applicants Accepted	• 65% • 30% • 5%
Faculty Resources	15%	• Ratio of 1998 Full-time Science and Clinical Faculty to Full-time Students	• 100%
Primary Care Rate	30%	• The Percentage of MDs From a School Entering Primary-care Residencies, Averaged Over 1996, 1997, and 1998	• 100%

<u>The National Research Council Ratings of Doctoral-research programs</u>

The 1982 (Jones, Lindzey, and Coggeshall) and 1995 (Goldberger, Maher, and Flattau) ratings of doctoral-research programs by the National Research Council (NRC) have been called the biggest, most expensive, most thoughtfully conceived and carried-out academic quality rankings ever done. Like the *US News* rankings, they are multidimensional. Unlike the *US News* rankings, they only rate doctoral-research programs, and they do not produce an overall ranking score.[7]

The 1982 study covered 2,699 programs in 32 disciplines or fields of study at 228 universities, and used between 12 and 16 indicators depending on the particular discipline being assessed. In its analysis of the results for each discipline, the committee showed the intercorrelations among all the indicators used. It also compared, for all disciplines included in each volume, the scores they received on the various indicators of quality. The committee did not, however, add together these scores in order to come up with an overall score nor did they rank the programs along any dimension. While this effort did not qualify as a proper ranking of programs, it laid the groundwork for the 1995 study.

The 1995 study was an update to the 1982 edition and contained information on 3,634 programs in 41 disciplines or fields of study at 274 universities. Information of two types was collected: objective measures of selected characteristics of doctoral-research programs, and the subjective views of faculty peers relative to program quality. The latter measure is similar to the reputational survey conducted by *US News*. Because of the elements of continuity with the 1982 study in terms of programs surveyed and indicators used, it was also possible for the committee to examine some of the changes that had taken place in higher education between the two studies.

7 One of the few studies that directly compared the *US News* and NRC rankings of doctoral programs was conducted in 1997 by Evan Rogers of Arizona State University and Sharon Rogers, a higher-education consultant in Virginia. The study concluded that the *US News* ratings of doctoral programs are as credible as those produced by the NRC. However, the doctoral program ratings form a small subset of *US News'* rankings and are generally not the focus of criticism. The study shed no light on the credibility of *US News'* graduate school rankings - i.e., the business, education, engineering, law, medicine, and primary-care rankings.

Whereas the 1982 report presented results by alphabetical listing of programs, the 1995 committee decided that rank ordering programs within fields would be a more convenient way for readers to review and interpret the information. The committee organized programs within each field into four groups or quarters based on the mean rating of the "scholarly quality of program faculty" and presented the rest of the information on these programs using the four groupings. They did not create an overall academic quality ranking, nor did they suggest that the mean rating of "scholarly quality of program faculty" could be used in this fashion.

The committee's reasons for avoiding the use of an overall ranking score are related to the way in which they viewed the nature of academic quality. In fact, the committee specifically identified the issues of unidimensionality, comparability, and freedom from error in their discussion of their methodology and results. Therefore, before discussing the reactions by various groups to the *US News* rankings and how these reactions relate to the issues of unidimensionality, comparability, and freedom from error, it is worth examining how the NRC dealt with these issues. While this discussion will serve as a context in which to view criticisms of the *US News* rankings, it should be remembered that the NRC rankings deal only with doctoral-research programs while the *US News* rankings cover both undergraduate and graduate programs.

How the National Research Council Committee on the Rating of Doctoral-Research Programs Dealt with the Issues of Unidimensionality, Comparability, and Freedom from Error

Unidimensionality. The summary of the 1995 NRC report on the rating of doctoral-research programs states: "Perhaps the most significant contribution this committee has made to the interpretation and use of the data ... is to present much of it in ranked order" (p. 60). However, while conceding to the superiority of ranking over other forms of data presentation in terms of its accessibility, the committee also points out the dangers of using a single number to describe any program and concludes that:

> it is not possible to provide a valid description of the quality of program by any method that relies exclusively on a single number. Rather than merely reporting where a given program ranks in its own field, it is critically important to indicate its relative standing on a number of measures. It is also important to report certain absolute quantitative measures of attributes that we believe are related to the quality of the education and training that the doctoral student receives at an institution." (Goldberger, Maher, and Flattau, 1995, p. viii, emphasis in original).

As previously mentioned, instead of weighing and summing the various indicators of academic quality to come up with an overall ranking score, the committee computed each institution's mean rating on the "scholarly quality of program faculty" indicator and used this to organize institutions within each field into four groups. The rest of the information was presented based on this organization. In case this approach would lead some readers to conclude that an institution's mean rating on "scholarly quality of program faculty" constituted its academic quality rank, the committee pointed out that:

> We have been particularly careful to incorporate a number of quantitative indices into our assessment variables, thereby placing reputational ratings [i.e., the "scholarly quality of program faculty" ratings] into a proper and modest perspective. In a word, there is no single agreed index of a unitary attribute called "quality"; there are several "qualities," and the importance of them is largely a function of the needs of the reader (p. ix).

It is interesting that the NRC's views on academic quality and how it can be measured result in rankings that do not comply with Webster's strict definition of an academic quality ranking. This also raises the question as to the validity of rankings that do meet Webster's definition.

Comparability. One of the main priorities for the committee was that they would be able to measure change in academic quality over time. Because of this, they tried to include as many doctoral programs as possible from the 1982 study.[8] In addition, they used several of the same indicators. Because of this and other elements of continuity between the 1982 and 1995 studies, it was possible to produce change measures in the following areas:

- Ratings of the "scholarly quality of program faculty"
- Number of faculty on a program-by-program basis
- Number of program graduates; and,
- Time to degree

[8] Of the 228 institutions in the 1982 study, 214 participated in the 1995 study. Over 1,900 doctoral programs in 27 fields appeared in both studies.

The committee found a remarkable degree of stability in all four categories. For example, 85% of participating programs that appeared in the top quarter for the "scholarly quality of program faculty" in 1982 appeared again in the top quarter in 1993.

The importance the NRC committee placed on maintaining continuity between studies is not evident in many other rankings. However, the value of this approach is apparent in the information it provides on changes in various aspects of program quality, both within and across programs, over time.

Freedom from error. The 1995 NRC committee raised several concerns about the reliability of the reputational measures used in the study. In particular, the committee noted that when the judgements of raters were pooled, there tended to be strong agreement about which programs were the strongest and weakest, but much less agreement about programs in the middle range. The committee went on to conclude:

> Because of the nature of reputational ratings, ... differences in ranked order between two programs may reflect very small, unreliable, or insignificant differences in the actual quality of a program and should be regarded by readers with great caution... Simple reputational rankings similar to those reported in the popular media may make for easier reading than the tables in this report. But because they mask subtleties that may be important to the reader, they also make for poorer information (p. ix).

The committee therefore included an appendix in the report that provided confidence intervals around each program's scholarly quality score. The overlapping confidence intervals for several programs indicated that there was statistically no (or little) difference in their quality ratings.

Overall, the NRC committee rejected the notion of a unidimensional attribute called academic quality that could be summed up in a single score, made comparability across years a priority, and acknowledged and illustrated the amount of error around the mean ratings for the "scholarly quality of program faculty." As mentioned previously, the NRC's stand on these issues is important to keep in mind as their work is viewed by many as a model for the rating of higher education programs.

Unidimensionality, Comparability, and Freedom from Error in the *US News* Rankings

Since their first appearance in 1983, responses to the *US News* rankings have been varied - from some suggesting that they are "among the best rankings ever published of any level of higher education" to others declaring that they are "depressing," "shallow and inaccurate," "built on quicksand," and "fundamentally dangerous" (Webster, 1992, p. 21). These responses are discussed here in terms of the extent to which they bear on the issues of unidimensionality, comparability, and freedom from error. *US News'* published views on these issues are also presented, where available.

Unidimensionality

When asked how the overall score on the *US News* rankings should be interpreted, a former *US News* editor replied that while the information used was multidimensional (covering categories such as Student Selectivity and Faculty Resources), the overall score was unidimensional, representing the overall academic quality of an institution (A.

Sanoff, personal communication, August, 1999). Many critics of the rankings (e.g., Crenshaw, 1999; Oslin, 1999) would disagree with this viewpoint and instead suggest that it is impossible to summarize a university's strengths in this fashion and that what is "best" in the eyes of *US News* may not be best in terms of the needs and desires of individual students (e.g., Ehrenberg, in Crenshaw, 1999; Levin, 1999).

As part of this argument, several critics have focused on problems with the particular indicators chosen (or not chosen) to represent aspects of academic quality. For example, some feel there are too few student-outcome indicators (e.g., Hicks, 1997; Seaman, 1998) while others (e.g., Wright, 1992) point out interpretation problems with some of the current indicators (e.g., SAT scores, and student/faculty ratios). In response to such criticisms, *US News* has acknowledged the need to include more student-outcome indicators but points out that the data are not being collected by any organizations and so are hard for them to access (A. Sanoff, personal communication, August, 1999). *US News* also points out that current indicators used were chosen for two reasons - either due to research identifying them as correlates of academic quality (no specific research is cited) or because they are of consumer interest. While the first selection method is a common in ranking studies, the second does not provide a basis for the use of these indicators in attempts to compare the academic quality of institutions.

Another part of the critics' argument focuses on the "one size fits all" nature of the indicators chosen to represent academic quality, pointing out that these indicators do not recognize the individual nature of schools in relation to their student populations and goals (Cantor, 1996; Lay, 1996) and that they oversimplify the college experience

(Smetanka, 1998). In the words of one college president, they are a process that "punishes us for filling our proper role in the community" (Cantor, 1996, p. 1). In particular, public institutions have complained that the indicators used (e.g., applicants accepted and graduation rate) are weighted against the larger, public schools (Cantor, 1996; Machung, 1998), pointing out that only three publics were ranked among the top universities in the country in the 1997 national university rankings, and that in the regional university lists, the publics were clustered in the second, third, and fourth tiers.

In addition to criticism of individual indicators used to represent quality, there are criticisms of the way in which they are combined to come up with a so-called academic quality score. An observation made by Webster (1986a) in relation to the National Research Council's decision not to provide overall rankings of doctoral-research programs in their 1982 report is interesting in terms of how it addresses this issue:

> [it is] easy to see the dilemma the committee faced in deciding how to present its data. It could scarcely have averaged the 12 to 16 measures it used for each discipline and listed schools by their rank on these, because for many or most of these measures their relation, if any, to academic quality is far from clear. If the committee had chosen the five or six of its measures that seemed most nearly to reflect academic quality, then it would implicitly have denigrated all those measures it had not used in its composite scores (pp. 122-123).

While *US News* avoids denigrating any indicators by using, for example at the college level, all 16 indicators to obtain an overall academic quality score, other media publications approach the composite issue differently. For example, *The Sunday Times* in England rated the quality of British universities using only six indicators: teaching quality, research quality, A-level points, employment, Firsts/2:1s awarded, and student/staff ratio (McCall, 1999). Interestingly, *The Sunday Times* also listed information for schools on several others indicators such as teaching staff, applications/places, and dropout rate that correspond closely to indicators used by *US News* but that were not incorporated into the rankings formula.

Certainly, there is no consensus on what constitutes academic quality and how it should be measured. In some ways, the popularity of the *US News* rankings - as evidenced by sales - would seem to indicate that the public buys into their model of academic quality. However, this may be less of a validation of the *US News* model, and more of an indication of the need that exists for comparable information on higher education institutions.

Comparability

An oft-cited criticism of the *US News* rankings is that annual changes in the weights and indicators used make it almost impossible to interpret shifts in a school's rank in terms of change in its relative academic quality (Levin, 1999; Machung, 1998; Pellegrini, 1999). For example, between 1993 and 1994, Georgetown fell eight places in the undergraduate rankings, from 17th to 25th. Apparently, the change was not due to a decline in Georgetown's academic quality but to a change in how one of the indicators in the Faculty Resources category was computed (Machung, 1998). A change in formula was also apparently behind the ranking flip-flop experienced by Johns Hopkins University, a well-established school, over the course of three years - from 21st to 10th to 15th (Cantor, 1996). These types of changes have been viewed as particularly unsettling since, as Machung (1998) points out, one of the most notable characteristics of national universities is that they actually do not change much from year to year and that they are characterized "much more by continuity than discontinuity" (p. 15).

Because *US News* does not provide much data on schools' performance on individual indicators, it is almost impossible to disentangle a change in rank produced by a change in formula from a change in rank produced by changes in a school's relative performance on the indicators. While some find this lack of year-to-year comparability a huge fault in the rankings, *US News* would argue that a trade-off exists and that it is preferable to make incremental changes every year to result in the "best possible rankings" than to use the same indicators every year to facilitate precise year-to-year comparisons (Webster, 1986b).

Freedom from Error

For many years, the *US News* rankings relied heavily on self-reporting (Smetanka, 1998). Critics have pointed out several error problems that arose due to mistakes (both accidental and deliberate) in reporting by institutions, and due to the differing ways in which schools compute figures for certain indicators (Machung, 1998). For example, a 1995 *Wall Street Journal* article (Stecklow, 1995) detailed a variety of misleading practices schools engage in to raise their SAT scores and graduation rates when reporting data to *US News* and college guidebook publishers. Wright (1990-91) also showed that, depending on which students were removed from the freshmen pool, the mean SAT score for Rochester University could move from 1149 (for the entire class) to 1218 (excluding minorities, athletes, legacies, foreigners and Not In Profile students). Her point was that schools are not uniform in which groups they retain when computing their mean SAT score and thus these figures in the rankings may not be comparable.[9]

US News has tried to reduce the error introduced by these practices by cross-checking data sent in by schools with data collected by debt-rating agencies, investors and national organizations such as the National Collegiate Athletic Association, and tightening up their survey questions (Stecklow, 1995). They also try to shame schools into reporting the correct data: for several years the magazine has listed law schools that have given them inaccurate LSAT scores. *US News* editor Robert Morse has also pointed out that whether or not schools fudge their data on some indicators probably has little effect on the final rankings since so many indicators are used that errors in a few of them will not significantly change the final outcome. In addition, he has pointed out that the reputational survey which accounts for 25% of the overall score would be very hard to tamper with (R. Morse, personal communication, August, 1999).

While reputational surveys may be hard to tamper with, they are not necessarily reliable (e.g., Goldberger, Maher, and Flattau, 1995; Wright, 1992). For example, some critics point out that error is introduced through low response rates (Smetanka, 1998). According to *US News*, this is not an issue for them as they survey all deans or relevant faculty in an area, not just a sample. The magazine points out that the NRC had to present confidence intervals around their reputation scores because they only sampled faculty on this measure (R. Morse, personal communication, August, 1999). While the

[9] Wright (1990-91) suggested using the SAT mid-range instead and *US News* took her advice on this.

response rate to *US News'* surveys can be as high as 70% (for the academic reputation survey for law schools in 1997), most surveys get a response rate in the 30 to 50% range. These response rates are below what is considered acceptable for surveys (Salant and Dillman, 1994), yet no adjustments are made. In fact, even when the response rate for the recruiters' reputational survey for business schools fell to an abysmally low 15% in 1995, no adjustment was made to schools' scores.

In addition to the above possible sources of error in the ranking data, there are other problems related to the nature of ranked information itself. Some observers note that one consequence of the conversion of tightly clustered interval-level data into ordinal ranks is to magnify small, and often insignificant, differences among schools (Machung, 1998). In Crenshaw's (1999) view, these "differences between schools that are close together are essentially meaningless. Small or even irrelevant changes by the school or the magazine can move a college half a dozen places up or down" (p. H01). As the Dean of the Law School at the University of Minnesota commented when his program rose from 20th to 17th in the 1998 rankings: "One can't really say with any precision, 'Oh, we're Number four, or Number 19,'.. At best we can say we're in a group" (Smetanka, 1998, p. 7B).

US News took a step towards dealing with this issue in 1998 when it began rounding overall scores to the nearest whole number. The magazine said that this adjustment was in recognition of the error that surrounds the overall scores and the fact that small differences after the decimal point may reflect non-significant differences between schools (Thompson and Morse, 1998, p. 66). Subsequently, the number of schools tied for overall score (and thus rank) increased dramatically.

Conclusion: The Future of Academic Quality Rankings

Whether or not the *US News* rankings indicators and methodology are a reasonable and accurate way of measuring academic quality, they influence students, parents, and institutions.[10] For example, Monk and Ehrenberg's (1999) empirical analysis of the ebb and flow of college applications over the past 15 years versus the ups and downs of various colleges in the *US News* rankings concluded that when a school moves up in the rankings, it attracts more and better applicants and obtains a higher "yield" (accepted applicants who come) from that crop. And when it moves down, the reverse happens. In fact, the study found that it is possible to look at a school's change in position in the rankings year to year, and predict just how much more (or less) selective it will be able to be the following year, how much its yield will improve (or decline) and how much the average SAT scores of incoming freshmen will rise (or fall).[11] At the

[10] However, the amount of influence varies. On the one hand, students themselves may not pay a lot of attention to the rankings. For example, one study found that only 10 percent of all college students pay serious attention to the rankings, another 30 percent pay them moderate attention, and a full 60 percent ignore them altogether (McDonough, Antonio, Walpole, and Perez, 1998). The same study also found that students who use college rankings are more likely to earn top grades, come from wealthy families, and be Asian-American. On the other hand, students' parents may pay more attention to the rankings, which in turn has an indirect effect on a student's choice of college. For example, Art and Science Group, a market research firm in Baltimore, Maryland, recently surveyed parents of high-achieving, college-bound students. Not only did the parents pay more attention to the rankings than their offspring, but a third cited *US News* as a "very reliable" source. Two-thirds found the rankings and other information provided by *US News* to be "very helpful" in evaluating a college's quality.

[11] The desire to do well on the rankings also influences some colleges' admissions policies (Machung, 1998; Stecklow, 1995). For example, the rise of early decision programs (students can apply to and receive an early decision from a college about five months ahead of the regular schedule) is seen by some as a strategy for colleges to improve their yield and attract more high ability students (i.e., those with high SAT scores).

graduate level, the influence of the rankings was evident in 1998 when the deans of all but 15 of the nation's 179 law schools signed a letter announcing "Law School Rankings May be Hazardous to Your Health!" that was sent to all those who took the LSAT in order to protest the impact that rankings have upon the choices of prospective law students (Caples, 1999; Cornwell, 1998).

Schools also realize that while rankings can hurt an institution, they can also help it in terms of self-promotion. In a survey of 160 presidents, provosts, and admissions deans at small private colleges, more than 92% said that the *America's Best Colleges* issue (*US News's* annual rankings of undergraduate colleges) does not paint an accurate picture of their schools. At the same time, 90% said the ratings are important - as a marketing tool (Editorial, The Christian Science Monitor, 1997). For example, when California State University, Chico was ranked one of the top regional universities in the West by *US News*, it incorporated this information into its marketing and recruitment campaigns (L. MacMichael, personal communication, September 27, 1999). Closer to home, a recent article in the *Boston College Chronicle* (Oslin, 1999) presented the college's 39th place ranking in the *US News* 2000 undergraduate rankings as a positive contribution to the college's image. In the words of the Dean of Enrollment Management: "The fact that we have reached the top-tier level of universities [in the *US News* rankings] continues to extend our reputation across the country" (p. 6). In addition, Boston College's standing on different indicators (e.g., graduation rate, applicant

---

Anecdotal evidence also indicates that admissions directors feel pressure to reduce their acceptance rate and do this by encouraging applications from students who they don't actually plan to admit in order to improve the applications-to-admittance ratio (Machung, 1998; Stecklow, 1995). Some schools, such as the acceptance rates, and faculty resources) was seen as a way of measuring the university's improvement over time: "We have improved in so many respects over the years...But we still need targets and benchmarks, and rankings are a way to remind us of how we compare against our peers" (p.6).

While rankings may have their uses, they are far from perfect. Needless to say, opinions on what constitutes the perfect academic quality rankings abound. Some of the suggested requirements include (Hattendorf, 1996; Webster, 1986a):

- Standardized reporting methods
- Multidimensional techniques
- Based on the achievements of most or all of a department's or university's faculty, students, and alumni
- Based on per capita, rather than aggregate figures
- Not only rank departments or institutions, but also show how they compare with some external standard of quality
- Be based partly on how much students learn

Some of these requirements have been met by various rankings efforts, but others are still unmet. For example, the rankings produced by the NRC and *US News* are multidimensional, use standardized reporting methods, and tend to use per capita rather than aggregate figures. However, these (and most other rankings) do not show how departments or institutions compare with some independent, external standard of quality, nor are they based on how much students learn.

There is also a growing consensus that the ideal academic quality rankings need to focus less on inputs and more on student-level outcome measures (Goldberger, Maher,

---

George Washington University School of Business and Public Management actually mount campaigns to and Flattau, 1995; Seaman, 1998). The current lack of student-outcome measures in rankings is partly due to the difficulty of identifying measures that can be quantified and compared across institutions. Another problem with identifying appropriate student-level outcomes is the fact that higher education itself is changing (Arenson, 1998).[12]

One of the more advanced - and intriguing - of the emerging student-outcome-based classifications is being developed by Robert Zemsky at the University of Pennsylvania (Seaman, 1998). Zemsky plans to produce by the year 2000 a kind of *Consumer Reports* on American colleges and universities that will rate schools on how graduates fare after they leave school. To test his methodology, Zemsky identified seven market segments and recruited fifteen institutions that spread across all seven. He had them survey their class of 1992 graduates six years out of college. Over the course of nine months, the participating schools were able to get a 48% response rate. Results for this pilot group indicate that what Zemsky calls "name-brand" schools produced more doctors and lawyers, while the "core" schools (a segment that includes most state universities) turned out more scientists and engineers, and "convenience" schools (which tend to sell education by the piece) turned out more teachers and nurses. Only name-brand schools sent a majority of graduates on to some form of further education. However, name brands did not necessarily lead the pack in graduates' income levels. While Zemsky's approach

---

recruit top students in order to improve their *US News* rank (Mufson, 1999).

[12] In recognition of this, the Carnegie Foundation, which developed the first college classification back in 1970, is overhauling its taxonomy to reflect the changes in many institutions. The explosion in the number of commuter colleges and, most recently, "virtual" universities that teach over the Internet poses a new dimension to these earlier classifications. Currently, only one out of every four college students in the US is an 18-to-21-year-old attending a traditional four-year college on a full-time basis.

is promising in terms of its emphasis on student-outcome indicators, doubts center on the validity of Zemsky's measures of such amorphous qualities as "confidence."

Rankings are also becoming more and more common at the elementary and high school levels (26 states now rank schools publicly), although here they are heavily based on one student-level outcome - standardized test scores (Cassidy, 1999; Finucane, 1999).[13] For example, in 1998, the Massachusetts State Department of Education announced that it would begin grading whole school districts based on scores on the state test, along with attendance, dropout rates, and other data; effectively placing schools in four or five tiers or categories such as excellent, improving, or poorly performing. In 1999, there was a proposal to take this a step further by using the information to rank the more than 2,000 schools statewide (Cassidy, 1999). The emphasis on test scores as a means of ranking schools was repeated at the national level when *US News* produced a report on outstanding American high schools in January, 1999 (Toch, 1999). In this instance, *US News* reviewed 1,053 schools in six metro areas and singled out 96 that they

[13] In a 1989 study, researchers studied the validity of various procedures for rank ordering school districts throughout a state on the basis of statewide test results. The various ranking procedures yielded widely varied ranks for most of the districts, regardless of whether scores were adjusted for different combinations of demographic factors. The conclusion was that any ranking procedure is likely to create inaccurate public and governmental perceptions of variation in educational programs, and may lead to misgiven conclusions regarding the programs' quality (Guskey and Kifer, 1989). Another study conducted at the state level (Robinson and Brandon, 1994) also concluded that average test scores should not be used to rank states according to the quality of their educational programs as most of the variation in scores could be explained by demographic factors over which states have no control. The report concludes that, even after attempting to adjust rankings by statistically adjusting the scores to reflect state variations in demographic characteristics, comparisons and rankings on the basis of these scores do not provide fair measures of educational quality.

felt could serve as models of excellence, based primarily on test scores and test-taking figures.[14]

Because of the increasing use of rankings to evaluate academic quality at all levels of the education system, it is more important than ever to understand the assumptions upon which various ranking methodologies are based, and to make sure that the rankings produced actually meet these assumptions. The issues of whether the score used to rank institutions actually represents academic quality, is comparable across years and is free from error should be investigated, particularly if the rankings are to be used for high stakes decision making about students or schools.

---

[14] *US News* created a "value added" model that measured a school's performance only after taking its students' family circumstances into account. The four measures of educational excellence used to evaluate schools were: state test scores, percentage of students taking the SAT or ACT, advanced placement test taking, and persistence rate. Excellent schools were those that had high persistence rates and performed better than expected on the three test-related measures. In other words, they added value.

Chapter Three

Methodology

Three techniques were employed to explore the technical issues surrounding *US News's* ranking methodology: factor analysis to explore the number of dimensions underlying the overall score; comparison tables to assess the degree to which the indicators and weights used to construct the rankings are comparable across years; and, the jackknife procedure to estimate the amount of error surrounding overall scores.

The same structure is used to discuss each technique. First, the technique itself is explained and any statistical tests involved are described. Second, the way in which the technique was applied to the *US News* rankings is outlined. This includes a discussion of the extent to which the data meet the underlying assumptions of the technique. Third, the format in which the results of the analysis are presented in Chapter Four is described.

Dimensions Underlying the Overall Rank Score

As discussed previously, the academic quality of an institution can never be measured directly. Instead, it must be measured indirectly through the use of indicators (e.g., institutional resources, student and faculty characteristics) that are theoretically related to the construct. Together, the information from these indicators can give a reasonably accurate "measure" of the institution's academic quality.

The accuracy of this measure will be affected by the extent to which important indicators of academic quality are omitted (e.g., the amount students learn during their time at college), and unrelated indicators are included (e.g., the number of vending machines on campus). Depending on whether one or both of these situations exist, the interpretation of an institution's academic quality score will be compromised. It is not always evident by just looking at the various indicators as to whether they are indirect measures of the same construct. Common sense and experience may suffice for judging their surface plausibility, but they do not provide quantitative evidence of the underlying construct(s) or dimension(s).

The Technique – Factor Analysis

Factor analysis is a technique used to identify the number of dimensions underpinning a set of indicators (Hair, Anderson, Tatham, & Black, 1998; Norusis, 1990a, 1990b). It begins by looking at the interrelationships or correlations among the indicators as a starting point for extracting the factors.[1] Each factor is expressed as a linear combination of the original set of indicators (Norusis, 1990a, 1990b). The factors represent the underlying dimensions that summarize or account for these indicators.

The uses of factor analysis are mainly exploratory or confirmatory depending on the objective of the researcher (Thompson and Daniel, 1996). An exploratory factor analysis is one that generally does not impose a solution at the outset but allows the technique to establish the relationship(s) and is generally appropriate when no previous studies have been conducted to clarify the number and types of dimensions underpinning

[1] A correlation coefficient indicates the strength of the linear relationship between two indicators. The value can range from –1 to +1, with +1 indicating a perfect positive relationship (both indicators get bigger

a set of indicators (Gorsuch, 1974; Johnson, Johnson, Gott, and Zimmerman, 1997; Nunnally, 1978). Confirmatory factor analysis is used to confirm a relationship or model in a set of indicators specified prior to the analysis and can be used to test rival models and to quantify the fit of each rival model (Hair et al., 1998). Exploratory factor analysis was used for the current study as no previous work exists on the number or type of dimensions that underpin the *US News* data, and also because of the extent of missing data, which makes it difficult to confirm any specific model. The rest of this discussion focuses on exploratory factor analysis.

Factor analysis usually proceeds in four stages (Norusis, 1990a, 1990b). First, the correlation matrix for all indicators is computed. A correlation matrix is a table showing the intercorrelations among all the indicators. Indicators that do not appear to be related to other indicators can be identified from the matrix and associated tests, and the appropriateness of using factor analysis can be evaluated.

Second, the factors are extracted from the data. Selection of the extraction method depends upon the researcher's objective. Principal components analysis is used when the objective is to summarize most of the original information in a minimum number of factors for prediction purposes. In contrast, common factor analysis is used primarily to identify underlying dimensions that reflect what the set of indicators have in common. While there is debate over which factor model is more appropriate (e.g., Mulaik, 1992), empirical research has demonstrated similar results in many instances

---

or smaller together), 0 indicating no relationship, and –1 indicating a perfect negative or reverse relationship (as one indicator gets bigger, the other gets smaller).

(e.g., Daniel, 1990; Gorsuch, 1974; Kim and Mueller, 1978). A principal components analysis was used in this study.

There are several rules for deciding when to stop factoring (that is, how many factors to extract). One of the more common crteria used (and the one that was used in this study) is the latent root criterion which focuses on the eigenvalues or latent roots that represent the amount of variance accounted for by a factor. According to this criterion, only factors having eigenvalues or latent roots greater than 1 are considered significant and retained (Kaiser, 1974).

After the factors have been extracted, the factor loadings can be examined. Factor loadings are the correlation of each indicator and the factor. Research has demonstrated that factor loadings have substantially larger errors than typical correlations and so should be evaluated at considerably stricter levels (e.g., ±.8 and above) in order to determine statistical significance (Gorsuch, 1974). Since very high loadings are hard to achieve, the practical significance of the loadings is a commonly used criterion, with loadings of ± .5 or greater considered practically significant.

The unrotated factor solution shown in the factor matrix may not provide a theoretically meaningful pattern of loadings. Thus, the third step in factor analysis is to rotate the factors in order to make them more interpretable. The major options available are orthogonal or oblique rotation methods. Orthogonal rotation means the axes are maintained at 90 degrees. This approach assumes the factors are not correlated. The oblique rotational method does not require that factor axes be orthogonal and is viewed as more realistic as the underlying dimensions are not assumed to be uncorrelated with each other. There are no specific rules to guide in selecting a particular orthogonal or oblique rotational technique. Instead, it is recommended that the choice of rotation technique be made on the basis of the particular needs of a given research problem (Gorsuch, 1974). Both rotations were employed for this study in order to assess the extent to which similar results would be obtained.

Once factors have been rotated and the factor loadings interpreted, they may be named. There is usually no single definitive (or best) factor solution for most problems. In this fourth stage of the factor-analytic process, factor scores can also be computed for each case (e.g., for each institution). A factor score is a composite measure created for each case on each factor extracted in the factor analysis. The factor weights are used in conjunction with the original indicator values to calculate each observation's score. The factor score then can be used to represent the factor(s) in subsequent analyses.

Replicating or relating factors. Sometimes, several factor analytic studies may be conducted in the same substantive area (e.g., factor analyses of the different academic quality rankings produced by *US News*). To determine whether a factor has been replicated, it is necessary to have objective means for relating factors from one study to those from another study (Gorsuch, 1974).[2]

According to Gorsuch (1974), two factors are related to the degree that they correlate together when scored for an appropriate set of individuals. In some situations,

[2] It is assumed that the factors from each study have been independently extracted and rotated. It is also assumed that factor hypothesis testing procedures are not appropriate. If confirmatory multiple group or

correlations are actually computed between the two sets of factor scores. In other situations, a subjective examination of the factor loadings of several studies may lead to conclusions as to the nature of the factors (e.g. Hirsch and Miller, 1991). The largest factor loadings for each factor can be underlined and compared across all studies to see if there is a consistent factor structure. While the latter approach was used in this study, it should be noted that the examination of factor loadings to relate factors has several problems. For example, it is only if the indicator loads .9 or so in both analyses that one can be absolutely certain that it is the same factor (Gorsuch, 1974).

Another way to assess the extent to which two factors are related is to examine the size of the factors. If the factors are actually the same, then the proportion of variance accounted for by each in the total matrix should be approximately the same in the two studies where the samples of indicators and of individuals have been drawn in the same manner. When oblique solutions are used, determination of the percent of variance accounted for by each factor must take into account both their direct and joint contributions. Even if factors differ in position and percent of variance accounted for, they may still be replicated across studies.

Significance tests. Exploratory factor analysis does not have many tests of statistical significance. It is heavily dependent on the researcher's knowledge in the theoretical area and his/her ability to interpret the factors.

---

maximum likelihood factor analysis can be computed, then it would be preferred over procedures for relating factors.

Some tests of statistical significance associated with factor-analytic techniques include the Bartlett Test of Sphericity and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy for evaluating the appropriateness of the correlation matrix for factor analysis; the chi-square (χ^2) significance test for judging the adequacy of the factor model;[3] and, the Stout Test of Unidimensionality (Stout, 1987) to determine whether a main factor underpins the set of indicators. The first two tests will be used for this study. While the fourth test also appears appropriate, it will not be applied as it is excessively data intensive.

Guidelines for the substantive or practical significance of a factor solution include the various criteria used to determine the number of factors to extract (e.g, the latent root criterion, which specifies that only factors with eigenvalues of 1 and above be retained) and the guidelines for the practical significance of the loadings for an individual factor (i.e., only loadings of ±.5 and above are considered significant).

The Application

In order to explore the number of dimensions underpinning the indicators used for the *US News* rankings, exploratory factor analytic techniques were applied to the 1999 and 2000 *US News* national university and national liberal arts colleges rankings at the undergraduate level, and to the 2000 business, education, engineering, law, medicine, and primary-care rankings at the graduate level. An SPSS computing package was used to

[3] The chi-square (χ^2) test associated with the maximum likelihood solution (a form of common factor analysis) can be used to test the adequacy of the factor model if the sample contains at least 51 more cases than the number of indicators under consideration (Lawley and Maxwell, 1971).

conduct the analyses. The extent to which the data meet the assumptions of the technique (which are outlined below) was checked prior to the analyses.

A principal components extraction method was used, followed by both orthogonal and oblique rotations. After the final factor solutions were obtained, factor loadings were interpreted within and across the various rankings.

Assumptions. The critical assumptions underlying factor analysis are more conceptual than statistical. One of the main conceptual assumptions is that some underlying structure exists in the set of selected variables. It is the responsibility of the researcher to ensure that the observed patterns are conceptually valid and appropriate to study with factor analysis. In terms of the current study, the ongoing debate over the extent to which the *US News* overall ranking score represents academic quality, and the extent to which the categories *US News* uses to produce sub-rankings are conceptually appropriate, would seem to warrant further investigation into the relationships among the indicators used.

From a statistical standpoint, the assumptions of normality and linearity apply only to the extent that they diminish the observed correlations (Hair et al., 1998). Normality is the degree to which the distribution of scores on an indicator corresponds to a normal distribution – i.e., the scores on the indicator should be clustered around the mean in a symmetrical, unimodal pattern known as the bell-shaped, or normal, curve. Lineanty refers to the concept that the model possesses the properties of additivity and homogeneity. In a simple sense, linear models predict values that fall in a straight line by having a constant unit change of the dependent variable for a constant unit change of the independent variable. The researcher must also ensure that the data matrix has sufficient correlations to justify the application of factor analysis. The Bartlett Test of Sphericity and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy are two statistical tests that can be used to assess the appropriateness of the correlation matrix for factor analysis.

In order to minimize the chances of "overfitting" the data, (i.e., deriving factors that are sample specific with little generalizability), the general rule for factor analysis is to have at least five times as many observations as there are indicators to be analyzed. In the case of the current study, only the schools of medicine and primary-care rankings did not have a sufficiently high cases-per-indicator ratio. While these were not excluded from the analyses, their results were interpreted cautiously. For some of the rankings (e.g., education and engineering), an acceptable cases-per-indicator ratio was maintained because not all of the original set of indicators used to compute the ranking score were available for factor analysis (see Tables 10, 11, 12, and 13 for a complete list of the indicators used in each ranking and their availability for this analysis).

It should also be noted that for most of the graduate school rankings (e.g., business, education, engineering, medicine, and primary-care), only data for the top 50 ranked institutions was available. This means that the relationships among indicators that were revealed through factor analysis may only apply to these institutions and not those ranked below.

Table 10
Indicator Availability for the 2000 Business and Law School Rankings[a]

Category	Business Indicators	Law Indicators
Reputation	Reputation Rank (Academic)	Reputation Score (Academic)
	Reputation Rank (Recruiters)	Reputation Score (Lawyers)
Selectivity	Average Undergraduate GPA	Undergraduate GPA (25/75)
	Average GMAT	LSAT (25/75)
	Acceptance Rate	Acceptance Rate
Placement Success	Median total Starting Compensation	Bar Passage Rate
	Employed at Graduation	Employed at Graduation
	Employed 3 Months After Graduation	Employed 9 Months After Graduation
Faculty Resources[b]	Not Applicable (N/A)	Student/Faculty Ratio
	N/A	**Average Expenditure Per Student for Financial Aid etc.**
	N/A	**Number of Volumes and Titles in the Law Library**
	N/A	**Expenditure Per Student for Instruction, Library, etc.**

[a]Shaded areas indicate data missing for this indicator
[b]This category of indicators was not used for the business school rankings

Table 11
Indicator Availability for the 2000 Education and Engineering Rankings[a]

Category	Education Indicators	Engineering Indicators
Reputation	Reputation Rank (Academics)	Reputation Rank (Academics)
	Reputation Rank (Superintendents)	Reputation Rank (Engineers/Recruiters)
Selectivity	Average Verbal GRE	Not Applicable (N/A)[b]
	Average Analytic GRE	Average Quantitative GRE
	Average Quantitative GRE	Average Analytic GRE
	Acceptance Rate	Acceptance Rate
Faculty Resources	Ph.D. Students/Faculty Ratio	Ph.D. Students/Faculty Ratio
	Ph.D's Granted	Ph.D's Granted
	M.A. Students/Faculty Ratio	**M.A. Students/Faculty Ratio**
	Masters Degrees Granted	N/A[b]
	Faculty Fellowships	Faculty Membership in the National Academy of Engineering
	Proportion of Graduate Students Who Were Doctoral Candidates	N/A[b]
	N/A[b]	**Proportion of Full-time Faculty Holding Doctoral Degrees**
Research Activity	Research Expenditure	Research Expenditure
	Research Expenditure Per Faculty Member	Research Expenditure Per Faculty Member

[a]Shaded areas indicate data missing for this indicator
[b]This indicator was not used for this particular ranking

Table 12
Indicator Availability for the 2000 Medical and Primary-care Rankings

Category	Medical Indicators	Primary-Care Indicators
Reputation	Reputation Rank (Academics)	Reputation Rank (Academics)
	Reputation Rank (Residency Directors)	Reputation Rank (Residency Directors)
Selectivity	Average Undergraduate GPA	Average Undergraduate GPA
	Average MCAT Score	Average MCAT Score
	Acceptance Rate	Acceptance Rate
Faculty Resources	Faculty/Student Ratio	Faculty/Student Ratio
Research Activity	National Institutes of Health Research Grants	Not Applicable (N/A)[a]
Primary-care Rate	N/A[a]	Percentage of Graduates Entering Primary-care Residencies

[a]This indicator was not used for this particular ranking

Table 13
Indicator Availability for 1999 and 2000 National University and National Liberal Arts College Rankings[a]

Category	Indicator	1999	2000
Academic Reputation	Academic Reputation Score	Yes	Yes
Retention	Freshmen Retention Rate	Yes	Yes
	Graduation Rate	Yes	Yes
Faculty Resources	% of Classes of 20 and Under	Yes	Yes
	% of Classes of 50 or More	Yes	Yes
	Faculty Salary	**No**	**No**
	Proportion of Professors With Highest Degree in Their Field	**No**	**No**
	Student/Faculty Ratio	Yes[b]	Yes[b]
	% of Faculty Who Are Full-time	No	Yes
Student Selectivity	SAT/ACT (25-75)	Yes	Yes
	Freshmen in Top 10% of High School Class	Yes	Yes
	Acceptance Rate	Yes	Yes
	Yield	**No**	**No**
Financial Resources	Financial Resources Rank[c]	Yes[b]	Yes[b]
Graduation Rate Performance	Graduation Rate Performance/Value Added	Yes	Yes
Alumni Giving Rate	Alumni Giving Rate	Yes	Yes

[a]Shaded areas indicate data missing for this indicator
[b]Data not available below top 50 schools
[c]This category had two indicators in 1999, and one in 2000. Only the rank is available for each year.

It should be acknowledged that this missing data could have had a significant effect on the estimation process and the factor structure that was revealed (Gorsuch, 1974). The degree of change in the solution is a function of the relationship of the missing indicator to the other indicators in the total set and the impact of that indicator as measured by the percent of variance it contributes to the solution (Gorsuch, 1974). The average amount of missing data in each ranking is approximately three indicators (representing about 18% of the original set of indicators). Since these missing indicators are generally from the Faculty Resources category (see Tables 10, 11, 12, and 13), there is some chance that a theoretically important factor may have been removed from, or obscured in, the final solution.

The Results of the Application

The results of the factor extractions are presented in table format in Chapter Four, as shown in Sample Table 1. A similar table is used to present the results for the undergraduate data.

Sample Table 1
Presentation of Factor Extraction Results for the Graduate School Rankings

Ranking (Number of Schools)	Number of Indicators Used	Number of Missing Indicators	Number of Factors Extracted	Percent Variance Accounted For – First Factor	Percent Variance Accounted For – Factor Solution
Business					
Education					
Engineering					
Law					
Medicine					
Primary Care					

The variance accounted for by the main factor in each solution is discussed in Chapter Four in terms of the extent to which it provides evidence of unidimensionality. In order for a measure to be considered unidimensional, the first factor extracted should account for the majority (e.g., 50 percent or more) of the variance. This is a factor with which most of the indicators are highly correlated. This does not necessarily mean that there is only one interpretable factor, but rather that there is a "large overriding factor with additional factors reflecting nuances of the factor structure" (Daniel, 1991, p. 10). The presence of a main factor is used by many researchers to provide support for the use of an overall score to summarize a case's performance on the construct (e.g., Camilli and Firestone, 2000; Hirsch and Miller, 1991; Powers and Gallas, 1980; Stout, 1987). If a unidimensional structure is not initially evident, subsequent analyses would be required to build an argument for the use of an overall score to summarize performance on the rankings indicators. For example, subsequent analyses may show moderate to high correlations among the different factors or dimensions, which would suggest that these dimensions are related in some way[4]. Scores on these dimensions could feasibly be added to describe a case's performance on the overall construct.

After interpreting and naming factors for each ranking, factor loadings for the following six pairs of rankings were compared in order to determine the extent to which factors may have replicated across different rankings or across different years of the same ranking: the 1999 and 2000 national university rankings; the 1999 and 2000 national liberal arts college rankings; the 2000 national university and national liberal arts college rankings; the 2000 business and law school rankings; the 2000 education and engineering rankings; and, the 2000 medicine and primary-care rankings. These pairs of rankings were chosen for comparison as they use very similar sets of indicators to obtain their overall scores.

Evidence of replication of factors *across rankings* may provide evidence of some consistency in terms of what *US News* is measuring, but not necessarily evidence that this something is academic quality. In addition, the comparisons *across years* at the undergraduate level (e.g., the comparison of the 1999 and 2000 national liberal arts college rankings) may provide evidence of the extent to which the factors obtained are generalizable across years, but not necessarily evidence that these factors reflect academic quality.

The Degree to which the Rankings are Comparable across Years

In general, if you want to assess the extent to which something remains constant over time (e.g., weight, temperature, achievement in mathematics), the device and unit of measure used to make the comparison should be the same each time (e.g., scale and pounds, temperature and Celcius, mathematics test and multiple-choice items). Beaton's aphorism, "when measuring change, do not change the measure" is the ideal (Beaton and Zwick 1992, p. 99).

---

[4] If the first-order factors are rotated obliquely, resulting in a matrix of correlations among the factors, it can be seen whether there are broader areas of generalization across the primary factors.

This standard would also seem appropriate for the realm of academic quality rankings. If you want to track an institution's academic quality over time, it is probably easier if you use the same measure(s) across time points. In this way, an institution's rise or fall in the ranks can be assessed in terms of their rise or fall on the same set of indicators. Not everyone would agree with this approach. For example, as discussed previously, *US News* views change in the indicators used to create their academic quality rankings as necessary if they are to improve in their ability to measure this construct and if they are to remain relevant (R. Morse, personal communication, August, 1999). An important question to ask, particularly in the context of the current debate over the annual changes in the *US News* ranking methodology, is how much change in the measure can be tolerated before the construct you are attempting to measure has shifted, or even changed entirely?

The Technique – Comparison Tables

Several different techniques may be used to quantify or illustrate the extent of change over time in a measure (Tukey, 1977). A particularly useful approach is that of graphical displays, which have been described as "instruments for reasoning about quantitative information" (Tufte, 1983, p. 9). Examples of simple graphical displays include pie charts and bar charts.

Pie charts can depict different types of information about a measure in the same visual space. The sections of the pie represent the different types of information being presented, and the size of each section represents its proportion of the total amount. For example, pie charts can be used to show the number and type of changes in ranking methodology on a year-by-year basis since these charts allow for both qualitative (i.e., changes in the definition of indicators) and quantitative (i.e., the number of indicators added to or removed from the rankings formula each year) changes to be presented in the same display.

Another useful display is a bar chart. The bar chart depicts a frequency distribution where frequencies are represented by bars. The length of the bars represents the number of cases (frequency) falling within each interval. Bar charts can be used to depict the total number of methodological changes for each ranking over time and to compare this total across rankings. In addition, bar charts can be used to depict the extent to which a constant set of indicators has remained embedded in the methodology for each ranking over time.

It is important to note that none of these graphic displays are meant to judge or evaluate the types of data they present. In addition, because these are descriptive techniques, there are no significance tests involved.

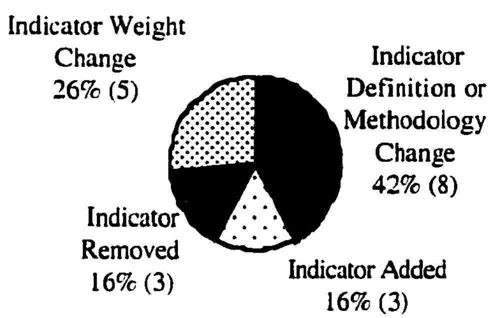
The Application and Results of the Application

Using pie and bar charts, changes over time in the indicators used to compute the rankings were quantified and graphically displayed. All of the data necessary to produce these charts were available from the *US News* guidebooks. The procedure was applied to the last five editions of the national liberal arts college, national university, business, education, engineering, law, medicine, and primary-care rankings.

The first stage of the comparison process focused on the types of changes that have occurred in ranking methodology over time for each ranking. The four types of changes tracked were: Indicator Weight Change, Indicator Removed, Indicator Added, Indicator Definition or Methodology Change. The basis for comparison in each case was the previous year's ranking methodology. For each type of change, the number of times it occured over the last five editions was summed. As shown in Sample Figure 1, the results for each ranking are presented in Chapter Four in the form of a pie chart that depicts the total number of changes occurring over the five-year period and the breakdown of these changes by type.

Sample Figure 1

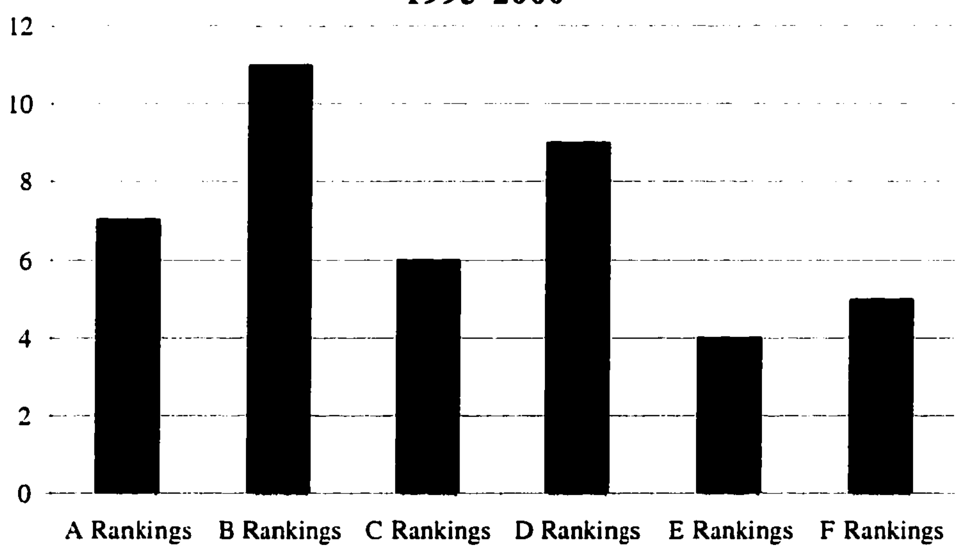
Types of Changes in Ranking Methodology



The next stage of the comparison process involved comparing the total number of changes in methodology over the last five editions of the guidebooks *across* rankings. As shown in Sample Figure 2, a bar chart is used to depict the total number of changes in each ranking's methodology. Each bar in the chart represents a different ranking – e.g., "A Rankings" could be the business school rankings, "B Rankings" the schools of education rankings, and so on. The height of the bar indicates the total number of changes over five editions in the methodology for that ranking. Because all rankings are included in the chart, it is possible to assess the extent to which changes in methodology have been occurring at the same rate across rankings.

Sample Figure 2

Total Number of Changes in Rankings Indicators, 1995-2000



The third stage of the comparison process involved comparing the extent to which a core set of indicators has remained embedded in the methodology for each ranking over the five-year period examined – in other words, the extent of non-change in the rankings. A bar chart is used to depict the proportion of unchanged indicators in each ranking. The discussion in Chapter Four focuses on the indicators that comprise this core set and the extent to which the core is comparable across rankings.

The fourth stage of the comparison process examined the extent to which the amount of changes in a ranking's indicators over the last five editions was related to the amount of movement in the relative ranks of schools over the same time period. This involved calculating the correlation between a school's rank in 1995 and 2000 for the graduate school rankings and between 1996 and 2000 for the undergraduate rankings. The results of this analysis shed further light on the link between changes in a ranking's formula and changes in schools' ranks.

The only assumption underlying these comparative analyses of the ranking methodology is that the types of changes occurring are discrete and can be summed in an additive fashion to give an indication of the overall extent of change from year to year.

The Amount of Error Surrounding Overall Scores

As previously mentioned, when trying to measure a construct such as academic quality, it is almost impossible to do so with complete accuracy. This is partly because we can only use indirect measures to "get at" academic quality. It is also due to problems with identifying, obtaining, or quantifying these indirect measures. If all relevant measures are not identified and included, specification error may result. This means that the set of measures specified to be indicative of academic quality may not be complete. In addition, if included measures contain inaccurate information due to mistakes in inputting data or calculations, measurement error will result. In the case of the *US News* rankings, specification error is hard to account for or quantify as (based on a review of the literature) there is no universally agreed upon set of measures for academic quality. Measurement error is easier to spot and *US News* is very careful to reduce or eliminate this type of error through checking and rechecking data for accuracy. Whether due to measurement or specification problems, when the various indicators of academic quality are combined, some error is inevitably contained in the overall score.

The Technique - Jackknifing

A *standard error* is a quantitative indication of the amount of uncertainty surrounding a single score. Instead of focusing on a single score for an observation, it makes us focus on a range of possible scores (e.g., the range of possible overall ranking scores for an institution). There are several different methods available for obtaining a standard error around a score, but here we will focus on the jackknife method. The jackknife method (Efron and Tibshirani, 1993; Quenouille, 1949; Tukey, 1958) starts with the complete data set and focuses on samples that leave out one observation at a time. Each time, an estimate of the standard error is computed. The jackknife estimate of standard error is computed using these multiple estimates.

The jackknife procedure can be coupled with a predictive procedure known as linear regression in order to obtain standard errors around scores (Gray and Schucany, 1972; Hinkley and Wei, 1984; Miller, 1964, 1974; Wu, 1989). A simple linear regression model involves a predictor (known as X or the independent variable) and a variable we are trying to predict (known as Y or the dependent variable). More complicated models involve multiple predictors and are used to analyze the relationship between a single dependent variable and several independent variables (e.g., between the overall ranking score and the indicators used to compute this score). The weights for the independent variables are chosen so that the correlation between the predicted and observed Y is maximal. The correlation between the predicted and observed Y is a measure of how well the best weighting of the independent variables predicts or correlates with the dependent variable. The higher the correlation between the dependent and independent variables, the better is the prediction; the lower the correlation, the greater is the margin of error in the predictions.

The jackknife procedure takes this basic model and reruns the regression analysis, removing one variable at a time (with replacement) and computing a predicted score for each case (e.g., each school in a ranking) each time. For example, if there are ten indicators in the basic regression model, the jackknife procedure will run the regression analysis ten times and each time one of the variables will be removed from the analysis before the regression model is computed. At the end of the ten runs, there will be ten predicted values for each school, one for each run. The jackknife standard error for a school is obtained from these values, using the following formula (Efron and Tibshirani, 1993):

$$\hat{se}_{\text{jackknife}} = \sqrt{\frac{n-1}{n}\sum\left(\hat{\theta}_{(i)} - \sum\frac{\hat{\theta}_{(i)}}{n}\right)^2}$$

where n is the number of regression models to be estimated and $\hat{\theta}_{(i)}$ is the predicted score for a school from the ith regression model with one indicator removed.

This jackknife standard error can be used in a statistical test known as a t-test to assess the extent to which one school's overall score is significantly different from that of another. The t-test formula employed for this study was:

$$t = \frac{x_1 - x_2}{\sqrt{\left(\hat{se}_{x_1}\right)^2 + \left(\hat{se}_{x_2}\right)^2}}$$

where x_1 is the overall score for school 1, x_2 is the overall score for school 2, $\left(\hat{se}_{x_1}\right)^2$ is the squared jackknife standard error for school 1, and $\left(\hat{se}_{x_2}\right)^2$ is the squared jackknife standard error for school 2.

Generally, the more t-test comparisons you make, the greater the probability of finding a significant difference between scores when there is none (known as a Type-I error). In order to control for this increased probability, the Bonferroni method for multiple comparisons can be used. This method uses the number of comparisons to be made as a way to determine the critical value of t (from the t-test) to be used for each comparison in order to keep the chances of making a Type-I error at a statistically acceptable level (for this study, .05). For example, if we want to keep the probability of a Type-I error at .05 and we are making 50 t-test comparisons, the probability of a Type-I error for each individual test would need to be kept at .001 and the critical t-value required for significance would be around 3.3.

Significance tests. The overall fit of the regression model can be assessed using several measures. A commonly used measure of goodness of fit is R^2. This coefficient can vary between 0 and 1. If the regression model is properly applied and estimated, one can assume that the higher the value of R^2, the greater the explanatory power of the regression equation, and, therefore, the better the prediction of the dependent variable (e.g., the overall score).

The sample R^2 tends to be an optimistic estimate of goodness of fit. The statistic *adjusted* R^2 attempts to correct R^2 to more closely reflect the goodness of fit of the model in the population. In terms of the current study, an *adjusted* R^2 of .9 and above for the overall model (i.e., before any indicators were removed as part of the jackknife procedure) was used as an indication that the model was doing a good job of predicting the dependent variable (e.g., the overall scores). The statistical significance of the R^2 and adjusted R^2 values can be assessed using an F test.

An F test can also be used to assess the relative importance of the independent variables used to predict the dependent variable. The importance of each variable is assessed by determining the significance of the change in R^2 when the variable is added to the equation. A large increase in R^2 indicates that a variable provides unique information about the dependent variable that is not available from the other independent variables in the equation.

The Application and Results of the Application

The jackknife procedure was used to obtain estimates of the amount of error surrounding an institution's overall score in the *US News* rankings. The sample of points to be jackknifed was the set of indicators used to compute the overall rank score for each institution. The analysis in Chapter Four focuses on the national liberal arts college and national university rankings at the undergraduate level and the business, education, and law school rankings at the graduate level.

Regression models were formed using SPSS. The extent to which the data meet the assumptions of the procedure was checked prior to, and after, conducting the analyses. All available indicators were used to construct the initial regression model for each ranking. In each case, the dependent variable was the overall score for schools. Although several methods are available for determining how independent variables are entered into the equation, a forced entry procedure was used for this analysis (all indicators entered together). The overall fit of the model to the data was assessed in terms of the *adjusted* R^2. Values of .9 and above were considered a good fit.

In the next stage of the procedure, one indicator in turn was removed from the analysis and the overall scores predicted using the remaining indicators. After each indicator in turn had been removed from the analysis (and replaced in all subsequent analyses), the jackknife standard error was obtained. This value was then used in a t-test to assess the extent to which one school's score was significantly different from that of another school. Since there are, on average, 50 schools in each ranking, around 49 t-test comparisons were made for each school in the rankings. In order to control for the increased probability of making a Type-I error due to the number of comparisons, a Bonferroni adjustment was applied.

Assumptions. Regression analysis and the jackknife procedure assume that the data are at the interval or ratio level. While most of the indicators used to produce the *US News* rankings are at the interval or ratio level, some are ordinal in nature (e.g., the reputation score). However, these ordinal indicators were still suitable for analysis as they approximate interval-level data due to the number of scale points involved.

The other assumptions of linearity, equality of variance, and normality for the regression model were checked by examining the residuals after the regression model had been built. In model building, a residual is what is left after the model is fit. It is the difference between an observed value and the value predicted by the model. For example, in order to check the linearity assumption, the predicted values for the dependent variable can be plotted on one axis and the observed values on the other. If the relationship is linear, the result should be a straight line. When evidence of violation of assumptions appears, you can either formulate an alternative model or you can transform the variables so that the current model will be more adequate (Glass and Hopkins, 1996).

It is important to note that the jackknife procedure assumes that the variables are randomly sampled from a population of variables. However, the indicators used to compute the rankings are more akin to a non-random convenience sample in that *US News* only picks indicators of academic quality that are easily quantified or for which data already exist and are easy to access. Because of this, the jackknife standard error produced is not strictly a standard error but more a general indication of the uncertainty surrounding the overall scores. For this study, the jackknife standard error represents the square root of the variance around a school's overall score due to changes in the indicators used to compute that overall score. In addition, this standard error is a conservative estimate of the error around scores as the indicators chosen by *US News* tend to be highly correlated. A random sampling from the population of indicators would probably be less highly correlated, which would result in larger standard errors around schools' overall scores.

The degree to which the *US News* data meet the assumptions of linear regression and the jackknife procedure may have been affected by the extent of missing data in the rankings. As previously discussed, data for entire indicators (e.g., the *Expenditure Per Student, Faculty Salary*, and *Yield* indicators) are missing in the public sources of ranking information made available by *US News* (see Tables 10, 11, 12, and 13). However, this type of missing data is not necessarily a problem in the construction of the regression model and the estimation of the jackknife standard error as long as there are sufficient indicators available to predict the overall score with a high degree of accuracy – i.e., if the *adjusted* R^2 is .9 or above, we can be fairly confident that the available indicators are sufficient for our task.

In addition to missing indicators, there are some instances of data missing for individual institutions on individual indicators and also some instances where indicators present a range of values rather than a single value. Several different methods can be used to fill in individual missing values. One of the simplest methods involves simply replacing missing scores with estimated values obtained from data on a highly correlated indicator. It has been pointed out that this method can distort the correlations among the indicators (Beaton, 1997). This possibility was assessed by comparing correlation matrices for the indicators prior to and after filling in the missing data. There are no set methods for dealing with indicators that use a range of values rather than a single value. Only two indicators of this type are used in the *US News* rankings – the 25^{th} to 75^{th} percentile for undergraduate SAT scores, and the 25^{th} to 75^{th} percentile for LSAT scores. In both instances, the value for the 25^{th} percentile test score was substituted in analyses as there is greater variance among schools at this lower end of the test score range. In addition, in the case of undergraduate schools that submitted ACT test score ranges, these were converted to SAT scores.

Chapter Four

Data Analysis and Results

This chapter presents the results of the analyses and discusses them in light of the issues of unidimensionality, comparability, and freedom from error. In preparation for the analyses, data were entered into an SPSS database and missing information filled in using estimated values derived from data on a highly correlated indicator. A random error term was also incorporated into the estimation. The effect of this estimation procedure on existing relationships among indicators was assessed by comparing the correlation matrices for the original and filled-in data sets. In every instance, there was little if any discrepancy.

While the amount of missing data in the *US News* rankings was generally small (amounting to a few schools per indicator), it was not missing at random - usually occurring in tiers three and four of the rankings. This was not an issue for the estimation of error or comparability analyses as these concentrated on schools in the first tier where there were fewer missing data. It was an issue for the factor analyses as these drew on all available data. While the effect of this non-random missing data on the factor analyses was probably small (since the amount of missing data was not extensive), it should still be kept in mind when evaluating these results.

Unidimensionality

The Analysis

Ten factor analyses were performed – one each for the 2000 business, education, engineering, law, medicine, and primary-care graduate rankings, and the 1999 and 2000 national liberal arts college and national university undergraduate rankings. In the first stage of each analysis, a correlation matrix was computed and results for the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and the Bartlett Test of Sphericity obtained in order to assess the suitability of the matrix for factor analysis. In every instance, the results for the two tests indicated suitability. The KMO varied between a "mediocre" .64 (for the education rankings data) and a "marvelous" .9 (for the law rankings data) and the Bartlett Test was highly significant (at the .001 level) in every case.

In the second stage of the analyses, factors were extracted using principal components (PC) analysis. Eigenvalue criteria were used to determine the number of factors to retain for the model. In the third stage of the analysis, the factors were rotated using both orthogonal and oblique rotation methods. The same factor structure was obtained after both rotations. Therefore, only the results for the orthogonal rotation are presented here and the oblique are presented in Appendix C. It should be noted that the correlations among the obliquely rotated factors are quite low – generally in the .1s and .2s - suggesting that the factors are not strongly related.

The Results

Tables 14 and 15 show the factor extraction results. The number of factors extracted varied between 1 for the law school rankings and 3 for the education, engineering and all undergraduate rankings (except for the 1999 national liberal arts college rankings). The first factor generally accounts for 50 percent or more of the variance, suggesting a unidimensional structure. While the first factors for education and engineering account for less than half of the variance (32.4 and 43.6 percent respectively) they are still quite large and suggest a dominant factor structure. The variance accounted for by the overall factor solution is large for all analyses, varying between 64.7 percent for law and over 80 percent for the 2000 national university rankings data. This suggests that the factor models are doing a good job of summarizing the relationships among the ranking indicators.

Table 14
Factor Extraction Results for the 2000 Graduate School Rankings

Ranking (Number of Schools)	Number of Indicators Used	Number of Missing Indicators	Number of Factors Extracted	Percent Variance Accounted For – First Factor	Percent Variance Accounted For – Factor Solution
Business (50)	8	0	2	59.7	76.9
Education (51)	10	4	3	32.4	67
Engineering (51)	10	2	3	43.6	76
Law (175)	9	3	1	64.7	64.7
Medicine (50)	7	0	2	55.1	71.5
Primary Care (51)	7	0	2	52.5	75.1

Table 15
Factor Extraction Results for the 1999 and 2000 Undergraduate Rankings

Ranking (Number of Schools)	Number of Indicators Used	Number of Missing Indicators	Number of Factors Extracted	Percent Variance Accounted For – First Factor	Percent Variance Accounted For – Factor Solution
National Liberal Arts 1999 (159)	11	6	3	49.4	74.9
National Liberal Arts 2000 (158)	11	5	2	52.7	68.3
National University 1999 (228)	11	6	3	51.3	79.4
National University 2000 (228)	11	5	3	52.3	80.5

While the rankings produced similar results in terms of variance accounted for, they use different numbers of indicators and schools and are subject to different amounts of missing indicators. The number of available and missing indicators most likely had an impact on the number of factors extracted, the variance accounted for by these factors, and the structure of the initial and rotated factors. For example, if data were available for the missing indicators in the education and engineering analyses (4 and 2 indicators missing respectively), the factor solutions would mostly likely change. While the business, medicine and primary-care rankings have no missing indicators, the small number (i.e., seven or eight) of available indicators resulted in rotated factor solutions that are not as substantially interpretable as some of the other solutions.

The number of schools for which data were available may also have significantly affected results. For example, only data on the top 50 ranked schools was available for each of the business, education, engineering, medicine, and primary-care factor analyses.

If data were available on the full population of schools in each case (varying between 124 and 317 schools), the relationships among indicators, and thus the final factor solutions, could change considerably. All of these data limitations point to the need for *US News* to release all data used to compile the rankings so that proper analyses of the indicators used can be conducted.

Tables 16 through 20 show the rotated factor structures for each graduate and undergraduate ranking. The rotated factor structures are presented in pairs in order to allow for the assessment of the extent to which rankings that use similar indicators have similar factor structures. These comparisons are somewhat hampered by the fact that rankings that use similar indicators may also have different indicators missing.

Table 16
Comparison of Rotated Factor Structures across Business and Law Rankings

Business Indicators (Loading)	Factor Name	Law Indicators (Loading)	Factor Name
Median Compensation (.86)* Average GMAT score (.85) Reputation Rank, Academics (-.84) Reputation Rank, Recruiters (-.82) Acceptance Rate (-.77) Average Undergraduate GPA (.74)	**Prestige**	Reputation Score, Lawyers (.91) Reputation Score, Academics (.90) LSAT (25th percentile) Score (.88) GPA (25th percentile) (.83) Percent Employed at Graduation (.64) Acceptance Rate (.58) Bar Passage Rate (.58) Percent Employed Nine Months After Graduation (.43) Student-faculty Ratio (.12)	**General Factor (Not rotated)**
Percent Employed at Graduation (.91) Percent Employed Three Months After Graduation (.87)	**Employment Success**		

*Loadings are rounded to two decimal places

Table 17
Comparison of Rotated Factor Structures across the Education and Engineering Rankings

Education Indicators (Loading)	Factor Name	Engineering Indicators (Loading)	Factor Name
Average GRE, Analytic (.87)	**Student Selectivity**	Number of PhDs Granted (.90)	**Reputation and Productivity**
Average GRE, Verbal (.82)		Reputation Rank, Engineers (-.88)	
Average GRE, Quantitative (.79)		Research Expenditure (.83)	
Acceptance Rate (-.61)		Reputation Rank, Academics (-.81)	
Reputation Rank, Superintendents (.90)	**Reputation**	Average GRE, Analytic (.84)	**Student Selectivity**
Reputation Rank, Academics (.88)		Average GRE, Quantitative (.82)	
PhD Student-Faculty Ratio (-.54)		Acceptance Rate (-.73)	
Number of Master's Degrees Granted (.84)	**Productivity**	Research Expenditure per Faculty Member (.84)	**Faculty Profile**
Number of PhDs Granted (.81)		PhD Student-Faculty Ratio (.70)	
Research Expenditure (.75)		Membership in the National Academy of Engineering (.69)	

Table 18
Comparison of Rotated Factor Structures across the Medicine and Primary-care Rankings

Medicine Indicators (Loading)	Factor Name	Primary-care Indicators (Loading)	Factor Name
Reputation Rank, Academics (-.89)	**Prestige**	Average MCAT Scores (.88)	**Student Selectivity**
Average MCAT Scores (.85)		Average GPA (.84)	
Research Grants (.83)		Student-Faculty Ratio (.83)	
Reputation Rank, Directors of Residency Programs (-.80)		Percent of Graduates Entering Primary-care Residencies (-.79)	
Average GPA (.77)			
Acceptance Rate (.94)	**Capacity**	Reputation Rank, Academics (.87)	**Reputation**
Student-Faculty Ratio (-.58)		Reputation Rank, Directors of Residency Programs (.80)	
		Acceptance Rate (-.62)	

Table 19
Comparison of Rotated Factor Structures Across 1999 and 2000 National Liberal Arts College Rankings

1999 Indicators (Loading)	Factor Name	2000 Indicators (Loading)	Factor Name
Academic Reputation (.92) SAT (25th percentile) Scores (.91) Top Ten Percent of High School Students (.85) Acceptance Rate (-.80) Freshmen Retention Rate (.80) Graduation Rate (.80) Alumni Giving Rate (.72) Percent of Full-time Faculty (.56)	**Student Inputs and Outputs**	Academic Reputation (.93) SAT (25th percentile) Scores (.92) Top Ten Percent of High School Students (.88) Acceptance Rate (-.81) Freshmen Retention Rate (.79) Graduation Rate (.77) Alumni Giving Rate (.73) Percent of Full-time Faculty (.58) Percent of Classes with Over 50 Students (.44)	**Student Inputs and Outputs**
Graduation Rate Performance (.91) Percent of Classes with Under 20 Students (-.76)	**Quality of Educational Experience (1)**	Graduation Rate Performance (.87) Percent of Classes with Under 20 Students (-.82)	**Quality of Educational Experience**
Percent of Classes with Over 50 Students (-.87)	**Quality of Educational Experience (2)**		

Table 20
Comparison of Rotated Factor Structures Across 1999 and 2000 National University Rankings

1999 Indicators (Loading)	Factor Name	2000 Indicators (Loading)	Factor Name
SAT (25th percentile) Scores (.94) Top Ten Percent of High School Students (.89) Academic Reputation (.87) Graduation Rate (.83) Acceptance Rate (-.83) Freshmen Retention Rate (.82) Alumni Giving Rate (.77)	**Student Inputs and Outputs**	SAT (25th percentile) Scores (.95) Top Ten Percent of High School Students (.90) Academic Reputation (.89) Graduation Rate (.85) Freshmen Retention Rate (.85) Acceptance Rate (-.81) Alumni Giving Rate (.80)	**Student Inputs and Outputs**
Percent of Classes with Over 50 Students (.88) Percent of Classes with Under 20 Students (-.74) Percent of Full-time Faculty (.71)	**Quality of Educational Experience**	Percent of Classes with Over 50 Students (.88) Percent of Full-time Faculty (-.73) Percent of Classes with Under 20 Students (-.73)	**Quality of Educational Experience**
Graduation Rate Performance (.96)	**Value**	Graduation Rate Performance (.97)	**Value**

For example, Table 16 shows the rotated factor structures for the business and law rankings. The first cell of the first column displays the names of indicators loading highly on the first factor for business schools. The loading or correlation is presented in parentheses after the indicator name. The higher the loading (whether positive or negative), the stronger the relationship between the indicator and the factor. The second cell of the first column lists indicators that load highly on the second factor, along with their loadings in parentheses. The second column in Table 16 summarizes the information in the first column by "naming" each factor according to the type of indicators loading high on it. While naming factors is a very subjective exercise, it helps in the identification of sub-constructs that underpin the overall academic quality score for a ranking. For example, indicators that describe the calibre of students accepted by business schools and the standing of each school in the eyes of corporate recruiters and academics all load highly (± .74 and above) on the first factor, which is characterized as a "Prestige" factor. The second factor has high loadings (.87 and .91) for indicators that measure one distinct outcome – employment rates – and is named "Employment Success". If schools were ranked separately on each of these factors or dimensions, a school might get a higher ranking in one area than another.

The third and fourth columns of Table 16 present the same type of information for law schools. While the business and law school rankings use fairly similar indicators, only one dimension appears to underpin the law school indicators. In addition, it is interesting that almost none of the categories of indicators that *US News* uses to produce sub-rankings for business and law schools (e.g., Reputation, Placement Success and Student Selectivity for business schools and Reputation, Placement Success, Student Selectivity, and Faculty Resources for law schools) match the dimensions obtained through factor analysis. Missing indicators may partly account for this outcome with the law school rankings, but not for the business school rankings as all indicators were available. However, the business school results may have been affected by the fact that data were only available for 50 of the 317 schools that *US News* compiled data on.

Table 17 shows the results for the education and engineering rankings, each of which have three underlying dimensions. While both have a "Student Selectivity" dimension, their other dimensions differ. For example, education has separate dimensions for indicators related to reputation and productivity, while engineering combines these into one dimension. One possible reason for this could be that, while the education and engineering rankings use similar indicators, these indicators may have different meanings and relationships across these schools. Another possible reason is that slightly different indicators are missing for each ranking and this is probably affecting the similarity of their factor solutions. It would be interesting to see how the factor solutions and structures would compare if the full compliment of information were available for both. (Some recommendations to this effect are outlined in Chapter Five.) It should also be noted that, as for the business and law schools, the obtained dimensions for schools of engineering (i.e., "Student Selectivity", "Reputation and Productivity", and "Faculty Profile") do not correspond to the categories that *US News* uses to produce sub-rankings for these schools (i.e., Reputation, Student Selectivity, Faculty Resources, and Research Activity). However, two of the three dimensions obtained for schools of education ("Selectivity" and "Reputation") are quite close to the categories that *US News* uses for these schools.

All indicators were available for the medicine and primary-care factor analyses, although this information was only available for the top 50 schools in each case. Both rankings have two underlying dimensions (see Table 18) that differ both from each other and from the categories *US News* uses to organize these indicators (i.e., Reputation, Student Selectivity, Faculty Resources, Research Activity [for the medical rankings only], and Primary-care Rate [for the primary-care rankings only]). The first medicine factor has high loadings for indicators that describe the calibre of students enrolled and the standing of the school in the eyes of medical academics and residency program directors and is named "Prestige." The second factor is a "Capacity" factor in that it has high loadings for indicators that describe the number of students accepted into a school and their ratio to faculty. The first primary-care factor is named "Student Selectivity" as it has high loadings for indicators describing student test scores (.88) and GPAs (.84). The second factor is a "Reputation" factor, which matches the category of indicators by this name used by *US News*.

The results of the factor analyses for the 1999 and 2000 national liberal arts college rankings indicate that they have very similar factor structures (Table 19). The first factor in each case is a "Student Inputs and Outputs" factor, with high loadings for indicators that describe the academic profile of incoming students and that measure their movement through the college experience. The second factor for both rankings is named "Quality of Educational Experience" as it has high loadings for indicators that describe how well served students were by their college education. The 1999 rankings produced a third factor and this loads only one indicator, *Percent of Classes with Over 50 Students*. This factor disappears in the 2000 rankings as the indicator loads instead on the "Student Inputs and Outputs" factor. The second and third factors for these rankings are weak in that they only load one or two indicators. If the full set of indicators were available, a more interpretable second and third factor might emerge. Similar to the other rankings, the underlying dimensions found in this analysis do not match the categories of indicators used by *US News* to create sub-rankings.

The results of the factor analyses for the 1999 and 2000 national university rankings are almost identical to each other (Table 20). Both have a first factor with high loadings for indicators that describe the calibre of entering students and their progress through college. This factor is named "Student Inputs and Outputs." The second factor is a "Quality of Educational Experience" factor with high loadings for indicators such as *Percent of Classes with Over 50 Students*, *Percent of Classes with Under 20 Students* and *Percent of Full-time Faculty* that describe the quality of resources and experiences available to students. Both third factors have very high loadings (.96 and .97 respectively) for the indicator *Graduation Rate Performance*, which also forms a category by the same name in the *US News* ranking methodology. These factor solutions exhibit several similarities to those obtained for national liberal arts colleges. For example, the first factors tend to have high loadings for indicators from the Academic Reputation, Retention, and Student Selectivity categories used by *US News* and the second factors draw on indicators from the Faculty Resources category. As for national liberal arts colleges, missing indicators come primarily from the Faculty and Financial Resources categories of indicators. The addition of these indicators to the analyses would most likely change the factor solutions for all the undergraduate rankings. In particular, they would probably produce a more substantially interpretable second factor for each of the rankings.

Overall, the results for the factor analyses are suggestive, but weakened by missing indicators in some cases, and small numbers of indicators in others. Both of these problems probably contributed to the lack of a substantially interpretable factor solution for certain rankings – e.g., the law, medicine, and primary-care solutions in particular. Despite these limitations, there are still some interesting findings. The fact that most of the rankings produced first factors that account for 50 percent or more of the variance provides strong support for the presence of a unidimensional structure in the indicators used to compute the academic quality scores. However, while this result suggests the presence of a dominant underlying attribute, it does not necessarily mean that the attribute is "academic quality". It just means that schools' performance on the indicators are strongly related and appear to be an indirect measure of some unseen attribute that has been named "academic quality" by *US News* but could actually represent another attribute or construct.

The results also indicate that usually two to three dimensions underpin a ranking and that these dimensions generally do not correspond to the categories *US News* uses to create sub-rankings. This suggests that if schools were ranked on the dimensions obtained through factor analysis, different sub-rankings (and overall rankings, depending on weighting) from those obtained using *US News'* categories might emerge.

The results of the comparative analyses of the factor structures for rankings that use similar sets of indicators are also suggestive, but weakened by missing indicators in some instances and small numbers of indicators in others. As Table 21 indicates, the most similar factor structures were obtained for the undergraduate rankings. At the graduate level, there is less similarity in factor structures, although certain individual factors (e.g., "Student Selectivity", "Reputation" and "Prestige" factors) tend to crop up across rankings. These results suggest that, while certain graduate rankings may use similar indicators (e.g., education and engineering), these indicators do not necessarily behave similarly across rankings. On the other hand, the results show a consistent factor structure across rankings and across years for the national liberal arts college and national university rankings.

Table 21
Named Factor Structures across Rankings

Business	Education	Engineering	Medical	Primary-care	1999 and 2000 National Liberal Arts	1999 and 2000 National Universities
Prestige			Prestige			
Employment Success						
	Student Selectivity	Student Selectivity		Student Selectivity		
	Reputation			Reputation		
	Productivity					
		Reputation and Productivity				
		Faculty Profile				
			Capacity			
					Student Inputs and Outputs	Student Inputs and Outputs
					Quality of Educational Experience	Quality of Educational Experience
						Value

Comparability

The Analysis

The first stages of the comparability analysis used simple counts and graphical representations to describe the nature and amount of change and non-change in the indicators used over the last five editions of the *US News* graduate (business, education, engineering, law, medicine, and primary care) and undergraduate (national liberal arts college and national university) rankings. Year-to-year comparisons (four comparisons in total, using the first edition as the baseline) started with the 1995 edition for graduate schools and the 1996 edition for undergraduate. The specific year-to-year changes for each ranking are documented in Appendix D. Appendix E contains a list of the core or unchanged indicators for each ranking during this time period.

Four types of changes were tracked: Indicator Weight Change, Indicator Removed, Indicator Added, and Indicator Definition or Methodology Change. Changes in weights, data collection methods, and the addition or removal of indicators were generally easy to track, although it was not possible to fully track changes in weights at the undergraduate level as this information was not included until the 1998 edition of the guidebook. Changes in indicator definition were harder to identify as the wording for a definition may differ from one year to the next, while the underlying meaning may not. The following rule was used to identify an indicator definition change:

(1) The new wording must contain additional detail such as a date, money amount, percent, or other precise information not previously alluded to or implied.

(2) If the new wording does not include such detail but includes phrases elaborating on or qualifying the previous year's definition, it should be recognized as changed by *US News* in the text of the guidebook

Since changes in an indicator's definition may lead to changes in the method of data collection and vice versa, they are subsumed under the same change category.

The last stage of the comparability analysis investigated the extent to which the number of changes in a ranking's indicators was related to the amount of change in the relative ranks of its schools over the last five editions. This involved calculating the correlation between a school's rank in 1995 and 2000 for the graduate school rankings, and 1996 and 2000 for the undergraduate, and comparing this to the extent of change in the indicators used to produce this rank.

The Results

Figures 1 through 7 present summaries of year-to-year changes in the indicators used for different rankings that are reflected in the last five editions of the *US News* graduate and undergraduate rankings guidebooks. Changes in the surveys *US News* use to collect data from schools are not reflected. Nor are changes in the way the overall rank score for a school is achieved. Changes in the latter occurred twice during the last five editions of the *US News* college and graduate school rankings: In 1998 when overall scores were rounded to the nearest whole number and in the 2000 edition when a school's score on each indicator was standardized before obtaining the overall rank score.

Figure 1

Changes in Law Indicators over Time

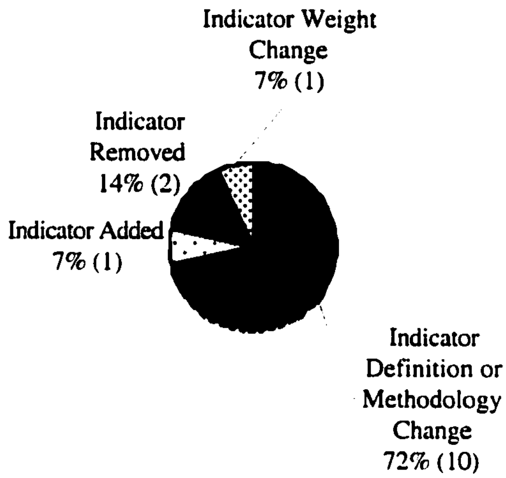


Figure 2

Changes in Business Indicators over Time

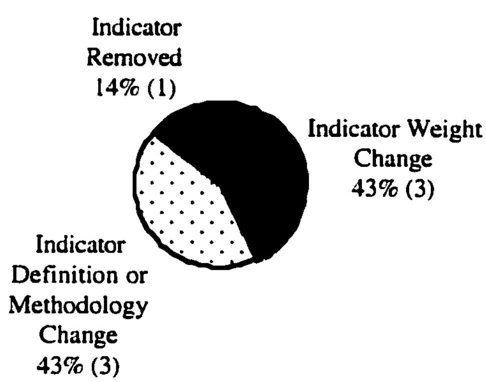


Figure 3

Changes in Education Indicators over Time

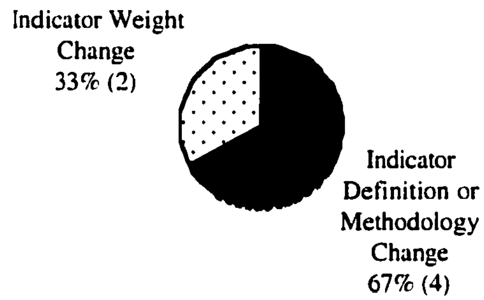


Figure 4

Changes In Engineering Indicators over Time

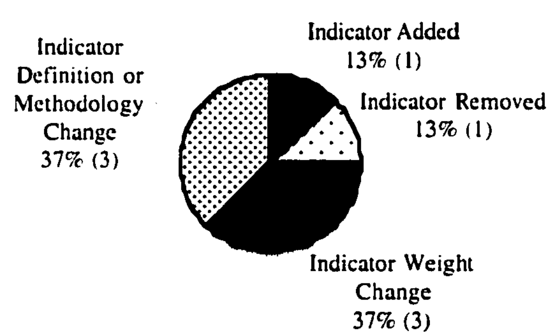


Figure 5

Changes in Primary-care Indicators over Time

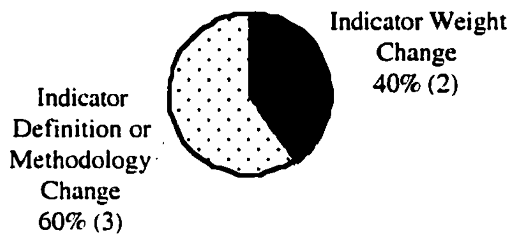


Figure 6

Changes in Graduate School Indicators over Time

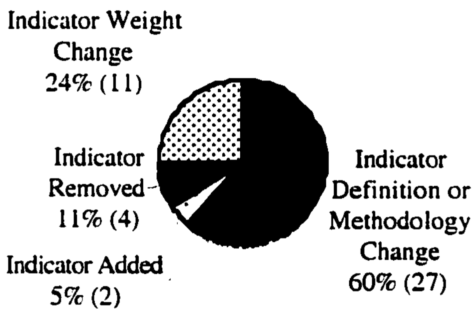
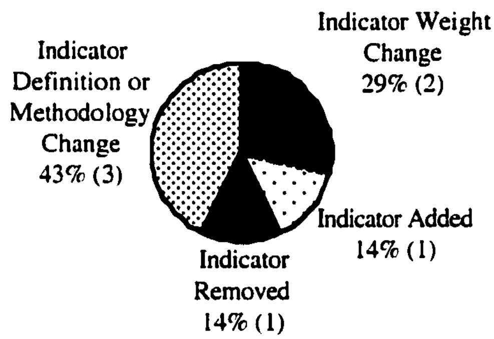


Figure 7

Changes in National University and National Liberal Arts College Indicators over Time



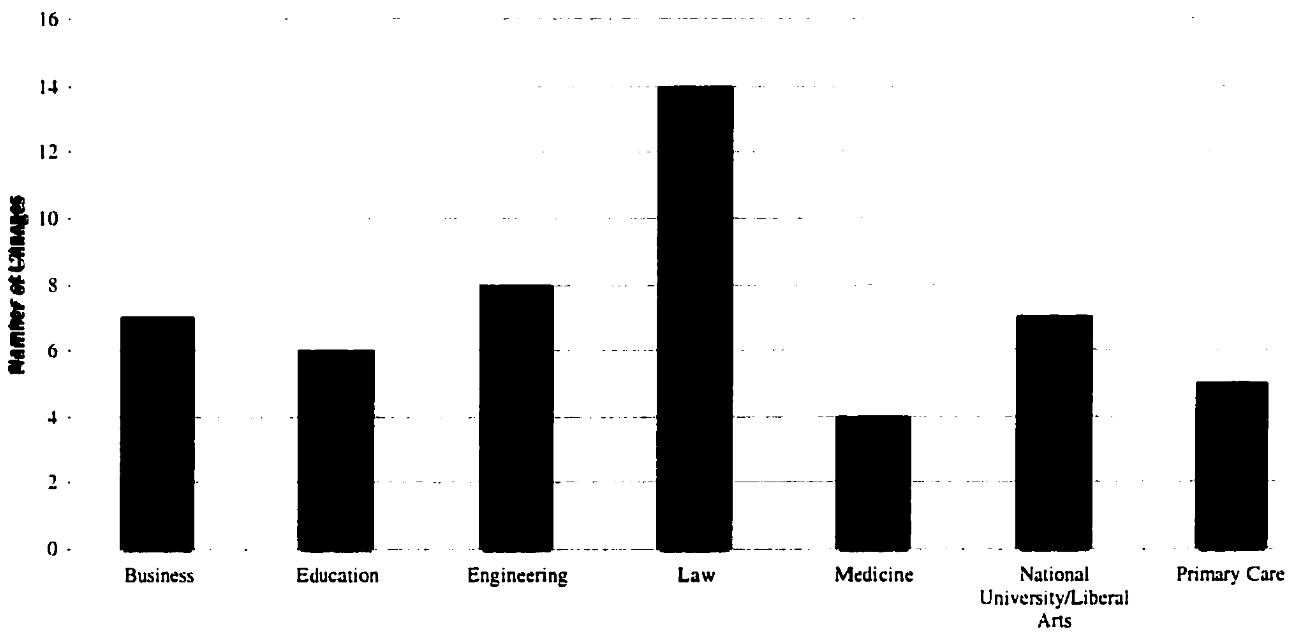
The law, engineering, and undergraduate rankings (Figures 1, 4, and 7) have experienced the most types of changes in their indicators although most of these changes are of two types – changes in an indicator's definition or methodology and changes in the weight assigned to a particular indicator. For example, since the 1995 edition, there have been 14 changes in the indicators used to rank law schools (Figure 1). Eleven (or 79 percent) of these changes are changes in the way an indicator is defined or measured or changes in the way it is weighted. The business, education, and primary-care rankings experienced fewer changes (Figures 2, 3, and 5), but these changes were also predominantly those of definition or weighting. For example, of the six changes occuring in the education indicators (see Figure 3), four were changes in definition or methodology and two were changes in weight. Changes in the indicators used for the medical school rankings are not displayed as there was only one type of change during this time period (four changes in the way indicators were defined).

Figure 6 summarizes the types of changes occurring across all graduate school rankings' indicators. Once again, we see that the majority (84 percent) of changes involve adjustments to what is already there – i.e., redefining an indicator or the way it is measured or adjusting its weight in the formula. Only 16 percent of changes involve adding to or removing from the existing formula. This pattern is consistent with that observed in the indicators used to construct the undergraduate rankings (Figure 7) where 72 percent of the changes (national universities and national liberal arts colleges are confounded here as they use the same indicators and formula) involve adjustments to the existing set of indicators. These results suggest that *US News* has generally retained the same set of indicators for each ranking, but has consistently refined and redefined these indicators over the years. (Of course, this redefining process can also change an indicator substantially). The cumulative effect of this process is discussed next.

In contrast to Figures 1 through 7, which focused on *type* of change, Figure 8 shows the *total number* of changes in the indicators for each ranking over the last five editions of *US News*.

Figure 8

Total Number of Changes in Indicators Used Over the Last Five Editions of *US News and World Report*

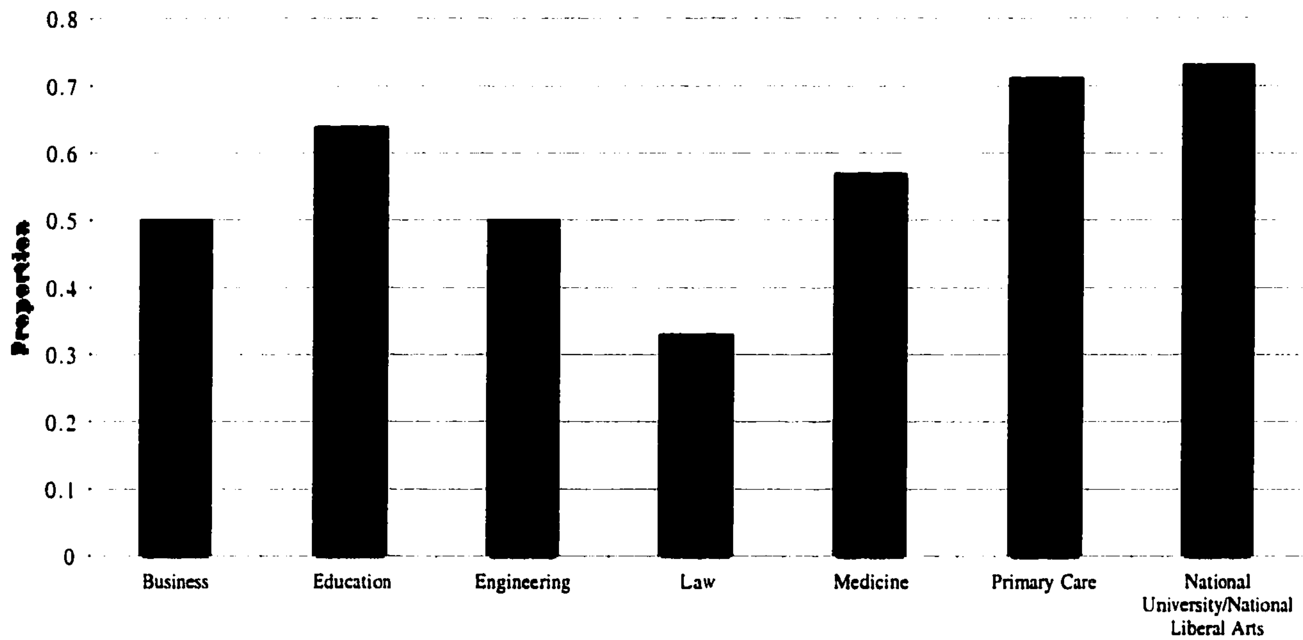


It is evident that the law rankings have experienced markedly more changes (14 in total) in their indicators than any other ranking, while the medicine and primary-care rankings have experienced the least (4 and 5 changes respectively). The rest of the rankings experienced between 6 and 8 changes in their indicators over this time period. Several reasons may account for the larger number of changes in the law ranking's indicators. One possible explanation is the amount of criticism that has been leveled against the law rankings, mainly by the schools themselves, and *US News's* response to that criticism. For example, the guidebook text accompanying the 1996 law rankings notes that "in response to suggestions from [law] schools, the magazine also refined its methodology, giving more weight to employment status six months after graduation and eliminating consideration of campus interview appointments per student" (p. 20). Another possible explanation is *US News'* response to the availability of new types of information on law schools. For example, in 1997, *US News* introduced a bar passage indicator into its law school ranking methodology in response to new information made available by an American Bar Association (ABA) survey of accredited institutions. It also changed the way it defined and measured student-faculty ratios and a school's success in job placement to be more in line with ABA guidelines in these areas. *US News* has made similar changes when new information has become available for other rankings (e.g., for the 2000 rankings, *US News* incorporated the MBA Career Services Council's versions of questions about job placement and starting salaries into the surveys sent to business schools). Despite these explanations, it is still interesting that the law schools have experienced twice as much change as other schools in their ranking formula. The extent of this change calls into question whether the law rankings are in any way comparable across years, and also raises the issue of how much change an academic quality ranking can tolerate and still represent the same construct.

While a ranking (such as the law school rankings) may have experienced a large number of changes relative to other rankings, these changes may be concentrated in a small group of indicators that are constantly being refined, and not spread out across the full set of indicators. Different rankings also use different numbers of indicators to compute their overall score and thus two rankings that experience the same types and number of changes may differ in the number of indicators left unchanged overall. Figure 9 shows the proportion of unchanged indicators for each of the rankings.

Figure 9

Proportion of Indicators Remaining Unchanged Over the Last Five Editions of *US News and World Report*



The undergraduate rankings (both national university and national liberal arts college) have the largest proportion (.73 approximately) of indicators that have remained unchanged. In contrast, only about one third of the law school ranking's indicators have remained unchanged over the last five editions of the guidebook. For most of the rankings, however, about half to two third of the indicators have remained unchanged during this time period. Appendix E shows the specific indicators for each ranking that have remained unchanged. Definite patterns emerge when comparisons are made across rankings. For example, most of the unchanged indicators are from two categories used by *US News* to organize its indicators: Student Selectivity and Faculty Resources. In the former category, test scores and the proportion of applicants accepted into the program are the most common unchanged measures of student selectivity. In the latter category, various forms of student-faculty ratios are the most common unchanged measures of institutional resources. The reason for this may be the ease with which test score, percent admitted, and student-faculty ratio data may be collected. In addition, some of these indicators – e.g., test scores – are strongly ingrained in the public's mind as symbols of institutional quality.

The final stage of the comparability analysis examined the extent to which the amount of change in indicators is related to the amount of movement in schools' ranks across years. Before presenting results in this area, it should be pointed out that *US News* cautions those attempting to interpret upward or downward movement in a school's rank from year to year since these changes can be the result of changes in formula as much as a change in quality.[1] Most of these warnings are aimed at prospective college students and not at the universities trying to understand their performance on the rankings over time. The results presented here are more focused on *patterns* in ranking performance over time and trying to understand how these patterns are similar or different across the different rankings and how they relate to the amount of change in a ranking's formula.

---

[1] See http://www.usnews.com/usnews/edu/college/rankings/cofaq.htm

Table 22 shows the correlations between a school's rank in 1995 and 2000 for each of the graduate rankings and between a school's rank in 1996 and 2000 for the undergraduate rankings.

Table 22
Correlation between 1995 and 2000 Rankings for Graduate Schools and 1996 and 2000 Rankings for Colleges

Ranking	**Correlation between Ranks**
Business	.91
Education	.76
Engineering	.91
Law	.91
Medicine	.88
National Universities	.97
National Liberal Arts College	.94
Primary Care	.07

There appears to be no definite relationship between the amount of change in indicators for a ranking and the correlation between its top 50 ranked schools in 1995 and 2000 at the graduate level and 1996 and 2000 at the undergraduate level. For example, while law schools experienced the most change in their indicators over the last five editions of *US News*, there was not much difference (r = .91) between the top 50 law schools in 1995 and in 2000. While varying amounts of change was experienced in the indicators used for the other rankings, they still show a high degree of similarity (approximately r = .9 and above) between their top 50 ranked schools in 1995 and 2000. The main exceptions to this are the education (r = .76) and primary-care (r = .07) rankings.

These results could be interpreted as meaning that, irrespective of the amount of change in indicators, schools are going to basically fall into the same order in terms of their academic quality. This is particularly evident in the top 20 ranked schools for every ranking – even for the education rankings – where most schools tend to hover around a certain rank. Change in a ranking's formula seems to produce more movement or re-ordering among schools at the lower end of the top 50. While this could be read as a validation of the *US News* rankings in terms of their ability to consistently identify the top schools in an area, it does not necessarily mean that these are the top schools in terms of academic quality (e.g., they may just be the most popular or well-known).

The low correlation between the primary-care ranks in 1995 and 2000 can be explained by changes in the population of schools that *US News* included in these rankings during this time period. For example, in 1995, the primary-care rankings were based on the 62 medical schools with the highest proportion of graduates between 1987 and 1989 entering primary care – i.e., pediatrics, general internal medicine and family practice. By the 2000 edition, all accredited medical schools (124 in total) and the 19 schools of osteopathic medicine were considered for these rankings. This extensive redefining of the population of schools used has not occurred for any of the other rankings.

In contrast, the low (relative to the other rankings) correlation between the 1995 and 2000 ranks for schools of education is linked to the fact that 16 of the top 50 schools in 1995 had experienced large changes in rank – of ten or more – by the 2000 edition. Table 23 shows the 16 schools of education that have experienced large changes in rank since 1995. The first five schools all experienced a decline in rank, ranging from a drop of 10 places for Florida State University and the University of Pennsylvania to a drop of at least 20 places (and out of the top 50 rankings) for Boston University. The remaining schools all improved their rank since 1995. Improvement ranged from an increase of 10 places for the University of North Carolina-Chapel Hill to a jump of 24 places for New York University.

It is interesting to note that there appears to be a certain momentum attached to these rises or falls in rank. Among the schools in Table 23 that began to fall after 1995, this tendency was generally sustained across each year. The same is evident among schools that began to rise after 1995. It would be interesting to see if and how schools that enter that upward or downward momentum eventually stabilize and what it takes to achieve that. For example, there may be a time dimension to improvement in certain areas such as test scores, their subsequent knock-on effect in areas such as retention and graduation rate and the cumulative effect of this on a school's upward or downward movement in the ranks.

Table 23
Schools of Education with the Biggest Differences in Rank between 1995 and 2000[a]

School	1995 Rank	1996 Rank	1997 Rank	1998 Rank	2000 Rank	Negative Change
Florida State University	30	35	36	37	40	-10
University of Pennsylvania	10	11	18	10	20	-10
Boston College	16	14	28	25	31	-15
Syracuse University	28	41	46	45	46	-18
Boston University	31	37	42	43	Not Rank ed	At least -20
School	**1995**	**1996**	**1997**	**1998**	**2000**	**Positive Change**
University of North Carolina-Chapel Hill	32	32	31	28	22	+10
George Washington University	45	39	37	30	34	+11
University of Oregon	28	30	20	37	16	+12
Cornell University	42	39	33	34	29	+13
Temple University	33	30	34	28	20	+13
University of Michigan-Ann Arbor	22	9	8	6	8	+14
University of Minnesota-Twin Cities	25	7	9	11	10	+15
Rutgers State University-New Brunswick	49	33	29	30	33	+16
University of Texas-Austin	27	19	12	13	11	+16
Arizona State University-Main Campus	47	29	39	27	24	+23
New York University	40	28	23	19	16	+24

[a]This table does not include schools that were not ranked in 1995 but appeared in the top 50 in the 2000 edition. There are six such schools: Washington State University, Washington University at St. Luis, University of California-Santa Barbra, Utah State University, SUNY Albany, and the University of Oklahoma.

Cross-year data for the top 50 schools in other rankings (i.e., business, engineering, law, medicine, national liberal arts college, national university, and primary-care) was also examined in order to assess the extent to which similar movements in rank occurred. Only eight business schools, three engineering, seven law, and three national universities differed by ten or more places in their 1995 and 2000 ranks. Until the 2000 edition, *US News* only printed the top 25 schools of medicine. Thus, only change in ranks for the top 25 schools could be tracked over the last five editions. For schools ranked in the top 25 in 1995, none showed a change in rank of 10 or more in the 2000 edition. *US News* only ranks the top 40 national liberal arts colleges. Only one national liberal arts college differed by ten places or more between the 1995 and 2000 editions (see Appendix F for more details on these analyses).

It is not clear why there was more movement in the relative academic quality of schools of education compared to that for other types of schools. Most of the schools in Table 23 experienced the biggest part of their upward or downward slide in rank between 1995 and 1996. However, a check of changes in the indicators used between 1995 and 1996 shows only one change in the definition of the *Percent of Faculty Given Awards* indicator. This indicator only accounts for 4% of the final overall score so its effect on any school's overall ranking score is bound to be slight. If changes in indicators are not responsible, movement must be due to change in schools' performance on the indicators. Ideally, it would be useful to track a school's performance on these indicators across years to see when and how change on an indicator may affect their overall rank. Of course, a school's change on an indicator or set of indicators needs to be considered in the context of how other schools are performing on these indicators. A school may improve on all indicators and still drop in rank if other schools also improve on these indicators. Unfortunately, it is hard to examine the effect of these indicators on a school's change in rank as *US News* did not print much information for these indicators until the

2000 edition. In addition, these analyses would still does not address the bigger question as to why there appears to have been more movement in the relative academic quality of schools of education compared to that for other types of schools.

Freedom from Error

The Analysis

For each of the 2000 business, education, law, national liberal arts college, and national university rankings, an estimation of the amount of error surrounding the overall score was produced using a regression model and the jackknife procedure. This error estimate was then used to determine whether a school's overall score in the *US News* rankings guidebooks differed significantly from that of another school's overall score. Since there were generally 50 schools in each analysis, each with their own standard error and overall score, each school had to be compared to 49 others (this varied across rankings). To adjust for the increased likelihood of a significant finding due to chance alone, a Bonferroni adjustment was applied when conducting the significance tests. The results of these tests were summarized in comparison tables and are discussed below. As discussed in Chapter Three, for this study, the jackknife standard error represents the square root of the variance around the overall score for a school due to changes in indicators used to compute this score.

The Results

The number of indicators available ranged between 8 for business and 13 for the national liberal arts college and national university rankings. This meant that the number of regression models estimated for each ranking varied between 8 and 13. In general, the percent of variance accounted for by the full regression model (i.e., when all available indicators were used to predict the overall score) was very high, varying between .98 for the business school, law school and national liberal arts college models, .94 for the national university model, and .91 for the education school model (all *adjusted* R^2 values).

The removal of one indicator at a time for the jackknife regression models did not seem to affect the overall *adjusted* R^2 in most instances. For example, for each of the 9 models estimated using the law school data (with one indicator removed each time), the *adjusted* R^2 never varied by more than .005 from the *adjusted* R^2 for the overall model (i.e., .98). This suggests that the indicators are contributing fairly similar information to the estimation of the overall score. Thus, when one is removed, the precision of estimation is not appreciably reduced. As a result, the jackknife standard errors for law schools are quite small, varying from a low of .82 for the University of Illinois-Urbana Champaign to a high of 3.56 for the University of Kentucky. Similar standard errors were obtained for all rankings except for schools of education (see Tables 24 through 28). The regression model for schools of education was not as robust to changes in indicators and the *adjusted* R^2 dropped considerably (by .14) when one indicator in particular – *Research Expenditure* – was removed. The resultant jackknife standard errors for schools of education are therefore quite large, varying from a low of 1.83 for Texas A&M up to 13.05 for UCLA (see Table 25).

Table 24
Standard Errors for Business Schools

School	Observed Score	Standard Error	School	Observed Score	Standard Error
Stanford	100.00	2.84	Georgetown	68	2.46
Harvard	98	1.31	Ohio State University	68	1.68
Northwestern	98	2.96	University of Maryland – College Park	68	3.31
University of Pennsylvania	98	2.07	University of Minnesota – Twin Cities	68	2.42
Massachusetts Institute of Technology	95	2.01	Washington Univetsity in St. Louis	67	1.72
University of Chicago	94	4.31	Arizona State University – Main Campus	66	3.76
Columbia University	93	1.90	Michigan State University	66	2.40
University of Michigan – Ann Arbor	93	1.48	Georgia Institute of Technology	65	2.74
Duke	91	1.38	University of Arizona	64	3.27
University of California – Los Angeles	90	1.97	University of California - Irvine	64	4.62
University of Virginia	89	2.49	Rice	63	2.72
Dartmouth	88	2.51	Southern Methodist University	63	1.59
New York University	86	0.91	Texas A& M University – College Station	63	2.32
University of California – Berkeley	84	1.46	Thunderbird Graduate School	63	2.80
Yale	83	1.52	University of California - Davis	63	4.54
Cornell	81	1.77	University of Illinois – Urbana-Champaign	63	2.13
University of North Carolina – Chapel Hill	81	2.37	University of Wisconsin-Madison	63	2.45
Carnegie Mellon	79	1.75	Wake Forest University	63	2.46
University of Texas – Austin	79	2.41	Case Western Reserve	62	3.25
Purdue	73	2.54	College of William and Mary	62	4.49
Emory	72	1.98	Tulane	62	2.56
Indiana University	72	1.35	University of Pittsburgh	62	3.19
University of Rochester	71	0.89	University of Washington	62	2.63
University of Southern California	71	2.76	BrighamYoung	61	3.88
Vanderbilt	69	1.51	Pennsylvania State University – University Park	61	1.63

Table 25
Standard Errors for Education Schools

School	Observed Score	Standard Error	School	Observed Score	Standard Error
Harvard	100	7.80	Pennsylvania State University – University Park	59	6.07
Teachers College	98	4.43	University of Iowa	59	3.73
Stanford	96	3.53	Cornell	58	7.94
University of California – Los Angeles	86	13.05	University of Colorado – Boulder	58	6.78
University of California – Berkeley	84	6.97	Boston College	57	2.82
Vanderbilt	82	9.86	University of Southern California	57	11.09
Ohio State University	80	5.16	Rutgers State University – New Brunswick	56	5.33
University of Michigan – Ann Arbor	78	2.44	George Washington University	55	6.51
University of Wisconsin-Madison	75	4.43	University of Missouri – Colombia	55	3.07
University of Minnesota – Twin Cities	71	7.27	University of Florida	54	3.08
University of Illinois – Urbana-Champaign	70	5.16	University of Pittsburgh	53	5.74
University of Texas – Austin	70	3.41	Washington State University	53	4.32
Indiana Bloomington	67	4.24	Washington University in St. Louis	52	8.41
Michigan State University	67	2.47	Florida State University	51	2.68
University of Virginia	67	3.80	University of Arizona	51	3.91
New York University	66	5.05	University of California – Santa Barbara	51	6.87
University of Oregon	66	9.07	Utah State University	51	7.11
Northwestern University	65	4.91	SUNY – Albany	50	6.13
University of Georgia	65	4.91	Texas A&M University – College Station	49	1.83
Temple	64	9.39	SUNY – Buffalo	47	5.78
University of Pennsylvania	64	8.18	Syracuse	47	5.36
University of Maryland – College Park	63	4.52	University of Connecticut	47	4.11
University of North Carolina – Chapel Hill	63	6.38	University of North Carolina – Greensboro	47	3.44
Arizona State University – Main Campus	60	6.01	University of Oklahoma	47	2.34
University of Kansas	60	7.72	University of Tennessee – Knoxville	47	4.73
University of Washington	60	3.17			

Table 26
Standard Errors for Law Schools

School	Observed Score	Standard Error	School	Observed Score	Standard Error
Yale	100	1.87	Emory	62	1.59
Harvard	93	2.56	BrighamYoung	61	2.49
Stanford	93	1.91	University of California – Hastings	61	2.38
New York Uni versity	89	2.16	University of Wisconsin-Madison	61	1.49
Columbia	87	1.03	Boston University	60	1.81
University of Chicago	86	1.99	College of William and Mary	60	1.53
University of Virgina	84	1.14	University of California – Davis	60	1.80
Duke	82	1.63	Washington University in St. Louis	60	1.96
University of Michigan – Ann Arbor	82	1.45	Indiana University – Bloomington	59	1.09
Cornell	81	1.28	University of Arizona	59	2.28
University of California – Berkeley	81	1.73	University of Georgia	59	1.59
Northwestern University	79	1.62	University of Utah	59	1.92
University of Pennsylvania	79	0.92	Fordham	58	3.35
Georgetown	77	1.22	Tulane	58	1.92
University of Texas - Austin	73	2.77	University of Connecticut	58	1.79
University of California – Los Angeles	72	3.56	Ohio State University	57	1.84
Vanderbilt	72	1.66	Wake Forest University	57	1.68
University of Minnesota – Twin Cities	70	1.94	University of Colorado – Boulder	56	1.98
University of Southern California	70	1.73	University of Florida	55	2.25
Washington and Lee University	67	2.07	University of Cinncinnati	54	2.90
University of North Carolina – Chapel Hill	66	2.04	University of Tennessee – Knoxville	54	0.94
University of Notre Dame	66	1.24	Arizona State University	53	1.97
University of Illinois – Urbana-Champaign	65	0.82	Baylor	52	2.94
University of Iowa	65	2.33	Southern Methodist	52	1.77
George Washington University	64	1.64	University of Alabama	52	2.09
University of Washington	64	1.94	University of Kentucky	52	3.56
Boston College	63	2.31			

Table 27
Standard Errors for National Liberal Arts Colleges

School	Observed Score	Standard Error	School	Observed Score	Standard Error
Swarthmore	100	1.53	Trinity	80	3.40
Amherst	99	2.12	Bates	79	2.90
Williams	97	1.74	Macalester	78	1.67
Wellesley	95	3.42	Barnard	77	2.93
Haverford	92	1.92	Colorado	77	3.16
Middlebury	92	2.80	Connecticut	77	1.17
Pomona	91	2.79	Oberlin	77	3.24
Carleton	90	3.23	University of the South	77	1.71
Bowdoin	89	1.20	Bucknell	76	1.25
Wesleyan	88	1.62	College of the Holy Cross	76	2.19
Davidson	87	2.64	Kenyon	74	1.54
Grinnell	87	1.60	Lafayette	73	1.50
Smith	86	2.67	Union	73	1.20
Claremont McKenna	85	2.90	Franklin and Marshall	72	3.08
Washington and Lee	85	2.76	Scripps	72	2.21
Mount Holyoke	84	1.73	Whitman	72	2.10
Vassar	83	1.40	Sarah Lawrence	70	2.67
Bryn Mawr	82	1.56	Dickinson	69	2.17
Colby	82	1.03	Bard	68	2.15
Colgate	82	1.40	Lawrence	68	2.45
Hamilton	82	1.25	Occidental	68	1.42

Table 28
Standard Errors for National Universities

School	Observed Score	Standard Error	School	Observed Score	Standard Error
California Institute of Technology	100	2.01	University of Michigan – Ann Arbor	73	2.37
Harvard	93	2.42	University of North Carolina –Chapel Hill	72	2.69
Massachusetts Institute of Technology	92	.73	Wake Forest	71	4.74
Princeton	91	2.80	College of William and Mary	70	1.95
Yale	91	1.85	Tufts	70	4.48
Stanford	89	1.92	Brandeis	69	1.47
Duke	86	1.39	University of California - San Diego	68	4.13
Johns Hopkins	86	1.25	University of Rochester	68	4.22
University of Pennsylvania	86	.88	Case Western Reserve	67	3.75
Columbia	85	1.23	Lehigh	67	2.54
Cornell	83	2.93	New York University	67	1.87
Dartmouth	83	.99	University of Illinois – Urbana-Champaign	67	1.80
University of Chicago	82	4.08	University of Wisconsin-Madison	67	4.41
Brown	81	2.92	Boston College	66	1.51
Northwestern	81	1.35	Georgia Institute of Technology	65	2.09
Rice	81	2.77	Pennsylvania State University	65	2.32
Washington University in St. Louis	80	1.49	University of California - Davis	64	1.84
Emory	78	1.93	University of Southern California	64	1.66
University of Notre Dame	77	2.28	Tulane	62	1.34
University of California - Berkeley	76	4.17	University of California - Santa Barbara	62	2.15
Vanderbilt	76	2.33	University of Texas - Austin	62	2.97
University of Virginia	75	.94	University of Washington	62	2.13
Carnegie	74	.91	Yeshiva	62	2.08
Georgetown	74	3.89	University of California – Irvine	61	2.23
University of California – Los Angeles	73	.99	University of Florida	61	1.06

The size of a school's standard error is not dependent on where they are in the rankings. This applies across all rankings. For example, Harvard, which is ranked first among schools of education (Table 25), has a much larger standard error (7.8) around its score than the schools ranked second and third (with standard errors of 4.43 and 3.53 respectively). Among the national universities (Table 28), twenty-second and twenty-third ranked University of Virginia and Carnegie Mellon are quite robust to changes in the indicators used to compute their overall scores, with standard errors of only .94 and .91 respectively. However, the University of Chicago which is ranked thirteenth is actually less robust to such changes, with a much larger standard error of 4.08 around its overall score.

Schools that have the same overall score do not necessarily have the same standard error either. For example, the six schools tying for forty-sixth rank among schools of education (Table 25) all have the same score (i.e., 47) but standard errors that vary between 3.44 and 5.78. This explains why schools that have the same overall score do not necessarily have the same pattern of results in the comparison tables. Differences in the standard errors for individual schools are due to differences among schools in terms of how the removal of different indicators from the equation affects the prediction of their overall score. For schools that have large standard errors, the removal of certain indicators from the estimation process makes it much harder to predict the overall score they received from *US News*. For school with small(er) standard errors, the removal of certain indicators from the prediction process does not reduce the precision of estimation of their overall score. This suggests that schools are differentially affected by the presence or absence of certain indicators in terms of their overall score and subsequent rank. In addition, the differences across rankings, in terms of the average size of the standard errors for schools in each ranking, suggest that rankings are differentially impacted by changes in their rankings indicators.

The results of the comparison analyses are summarized in Tables 29 through 33. In each table, schools are ordered by their overall ranking score across the heading and down the rows. Read across the row for a school in order to compare its performance with the schools listed in the heading of the chart. The symbols indicate whether the overall score of the school in the row is significantly lower than that of the comparison school in the heading (arrow pointing down), significantly higher than that of the comparison school (arrow pointing up), or if there is no statistically significant difference between the two schools (circle). The blank diagonal represents where a school is compared against itself.

If there was no error around the overall scores for schools, Tables 29 through 33 would only consist of arrows pointing up and down, except for instances where two schools have the same overall score and are tied for rank. This is not the case. For example, in the business school rankings comparison table (Table 29) Stanford is listed first in the row and heading as it has the highest overall score among business schools. However, reading across the row, it appears that Stanford's overall score of 100 is not significantly different from that of twelve other schools that are ranked beneath it. These 12 schools include Harvard, ranked second with a score of 98, and New York University,

Table 29
Multiple Comparisons of Overall Ranking Scores - Business Schools

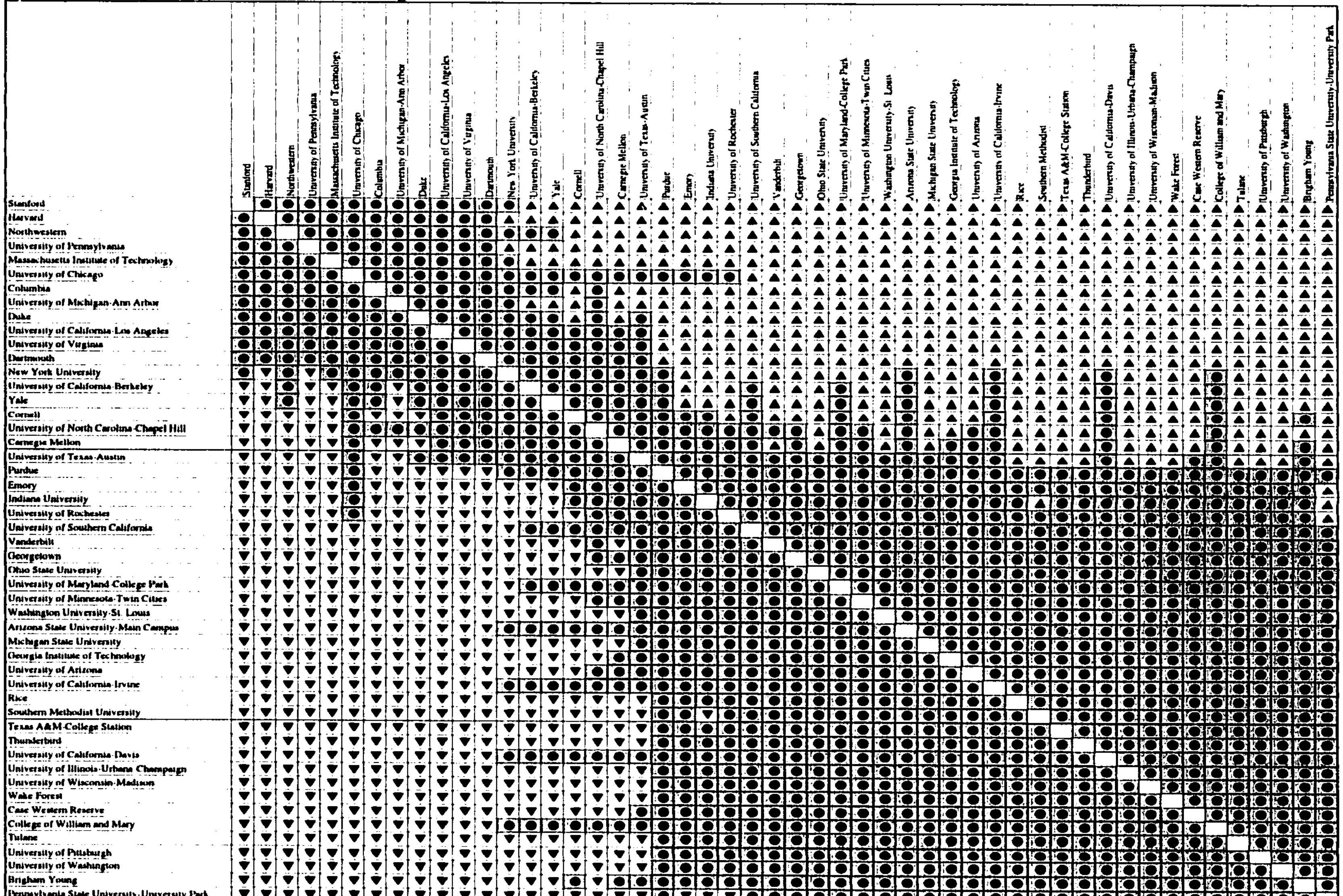


Table 30
Multiple Comparisons of Overall Ranking Scores - Schools of Education

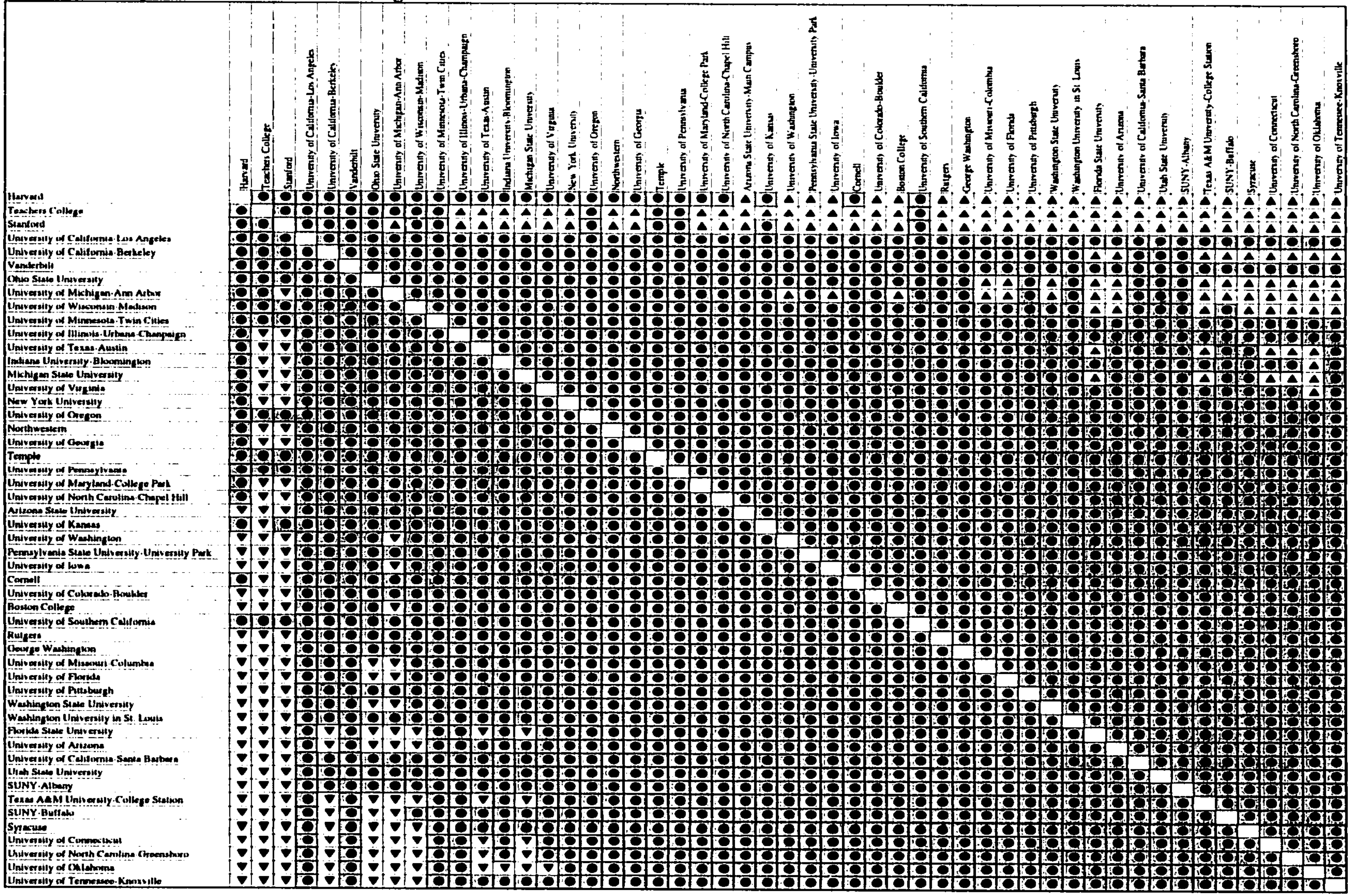


Table 31
Multiple Comparisons of Overall Ranking Scores - Law Schools

	Yale	Harvard	Stanford	New York University	Columbia	University of Chicago	University of Virginia	Duke	University of Michigan-Ann Arbor	Cornell	University of California-Berkeley	Northwestern	University of Pennsylvania	Georgetown	University of Texas-Austin	University of California-Los Angeles	Vanderbilt	University of Minnesota-Twin Cities	University of Southern California	Washington and Lee University	University of North Carolina-Chapel Hill	University of Notre Dame	University of Illinois-Urbana-Champaign	University of Iowa	George Washington University	University of Washington	Boston College	Emory	Brigham Young	University of California-Hastings	University of Wisconsin-Madison	Boston University	College of William and Mary	University of California-Davis	Washington University in St. Louis	Indiana University-Bloomington	University of Arizona	University of Georgia	University of Utah	Fordham	Tulane	University of Connecticut	Ohio State University	Wake Forest	University of Colorado-Boulder	University of Florida	University of Cincinnati	University of Tennessee-Knoxville	Arizona State University	Baylor	Southern Methodist	University of Alabama	University of Kentucky
Yale		●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Harvard	●		●	●	●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Stanford	●	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
New York University	●	●	●		●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Columbia	▼	●	●	●		●	●	●	●	●	●	●	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
University of Chicago	▼	●	●	●	●		●	●	●	●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
University of Virginia	▼	●	●	●	●	●		●	●	●	●	●	●	▲	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Duke	▼	●	▼	●	●	●	●		●	●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
University of Michigan Ann Arbor	▼	●	▼	●	●	●	●	●		●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Cornell	▼	●	▼	●	●	●	●	●	●		●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
University of California-Berkeley	▼	●	▼	●	●	●	●	●	●	●		●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Northwestern	▼	▼	▼	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
University of Pennsylvania	▼	▼	▼	●	▼	●	●	●	●	●	●	●		●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Georgetown	▼	▼	▼	▼	▼	●	▼	●	●	●	●	●	●		●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
University of Texas-Austin	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
University of California-Los Angeles	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	▲	▲	▲	●
Vanderbilt	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	▲	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
University of Minnesota-Twin Cities	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	▲	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
University of Southern California	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	●	▲	●	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Washington and Lee University	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	▲	▲	▲	●
University of North Carolina-Chapel Hill	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	●
University of Notre Dame	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	●	●	●	●	●	●	▲	▲	●	●	▲	▲	●	▲	▲	●
University of Illinois-Urbana-Champaign	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	▲	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	●
University of Iowa	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●
George Washington University	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	●
University of Washington	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	●	▲	▲	●
Boston College	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Emory	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	▼	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	●	▲	●	●
Brigham Young	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
University of California-Hastings	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
University of Wisconsin-Madison	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Boston University	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
College of William and Mary	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
University of California-Davis	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Washington University in St. Louis	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Indiana University-Bloomington	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	●	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
University of Arizona	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
University of Georgia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
University of Utah	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●
Fordham	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●
Tulane	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
University of Connecticut	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●
Ohio State University	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●
Wake Forest	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●
University of Colorado-Boulder	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●
University of Florida	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●
University of Cincinnati	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●
University of Tennessee-Knoxville	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●
Arizona State University	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●
Baylor	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●
Southern Methodist	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●
University of Alabama	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●
University of Kentucky	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	

Table 32
Multiple Comparisons of Overall Ranking Scores - National Liberal Arts Colleges

	Swarthmore	Amherst	Williams	Wellesley	Haverford	Middlebury	Pomona	Carleton	Bowdoin	Wesleyan	Davidson	Grinnell	Smith	Claremont McKenna	Washington and Lee	Mount Holyoke	Vassar	Bryn Mawr	Colby	Colgate	Hamilton	Trinity	Bates	Macalester	Barnard	Colorado	Connecticut	Oberlin	University of the South	Bucknell	College of the Holy Cross	Kenyon	Lafayette	Union	Franklin and Marshall	Scripps	Whitman	Sarah Lawrence	Dickinson	Bard	Lawrence	Occidental
Swarthmore		●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Amherst	●		●	●	●	●	●	●	▲	▲	●	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Williams	●	●		●	●	●	●	●	▲	●	●	▲	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Wellesley	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Haverford	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	▲	▲	●	▲	▲	▲	●	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Middlebury	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	▲	●	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Pomona	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Carleton	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Bowdoin	▼	▼	▼	●	●	●	●	●		●	●	●	●	●	●	●	●	●	▲	●	▲	●	●	▲	▲	●	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Wesleyan	▼	▼	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	●	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Davidson	▼	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Grinnell	▼	▼	▼	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Smith	▼	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	▲	▲	▲	▲	▲
Claremont McKenna	▼	▼	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	▲	▲
Washington and Lee	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	▲	▲	▲
Mount Holyoke	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲
Vassar	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲
Bryn Mawr	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	●	●	●	▲	▲	▲	▲
Colby	▼	▼	▼	▼	▼	●	●	●	▼	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	▲	●	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲
Colgate	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	●	●	●	▲	▲	▲	▲
Hamilton	▼	▼	▼	▼	▼	●	●	●	▼	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲
Trinity	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Bates	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲
Macalester	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲
Barnard	▼	▼	▼	▼	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Colorado	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Connecticut	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲
Oberlin	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●
University of the Sou	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	▲
Bucknell	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	▼	●	●	●	●	●	●	▼	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	▲
College of the Holy C	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●
Kenyon	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	▼	▼	●	▼	●	▼	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●
Lafayette	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●
Union	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●
Franklin and Marshal	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●
Scripps	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	▼	▼	●	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●
Whitman	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	▼	▼	●	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●
Sarah Lawrence	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●
Dickinson	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●
Bard	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●
Lawrence	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●
Occidental	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	●	●	▼	●	▼	▼	●	●	●	●	●	●	●	●	●	●	●	

Table 33
Multiple Comparisons of Overall Ranking Scores - National Universities

	California Institute of Technology	Harvard	Massachusetts Institute of Technology	Princeton	Yale	Stanford	Duke	Johns Hopkins	University of Pennsylvania	Columbia	Cornell	Dartmouth	University of Chicago	Brown	Northwestern	Rice	Washington University in St. Louis	Emory	University of Notre Dame	University of California-Berkeley	Vanderbilt	University of Virginia	Carnegie Mellon	Georgetown	University of California-Los Angeles	University of Michigan-Ann Arbor	University of North Carolina-Chapel Hill	Wake Forest	College of William and Mary	Tufts	Brandeis	University of California-San Diego	University of Rochester	Case Western Reserve	Lehigh	New York University	University of Illinois-Urbana-Champaign	University of Wisconsin-Madison	Boston College	Georgia Institute of Technology	Pennsylvania State University	University of California-Davis	University of Southern California	Tulane	University of California-Santa Barbara	University of Texas-Austin	University of Washington	Yeshiva	University of California-Irvine	University of Florida
California Institute of Technology		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Harvard	●		●	●	●	●	●	●	●	●	●	●	●	●	▲	●	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Massachusetts Institute of Technology	●	●		●	●	●	●	▲	▲	▲	●	▲	●	●	▲	●	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Princeton	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	▲	●	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Yale	●	●	●	●		●	●	●	●	●	●	●	●	●	▲	●	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Stanford	▼	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	▲	●	▲	▲	▲	●	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Duke	▼	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	▲	●	▲	●	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Johns Hopkins	▼	●	▼	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	▲	●	▲	●	▲	▲	●	▲	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
University of Pennsylvania	▼	●	▼	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	▲	●	▲	●	▲	▲	●	▲	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Columbia	▼	●	▼	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	▲	●	▲	●	▲	●	●	▲	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Cornell	▼	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	●	●	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Dartmouth	▼	●	▼	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	▲	▲	●	▲	●	●	●	▲	●	▲	●	●	●	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
University of Chicago	▼	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Brown	▼	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Northwestern	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	▲	●	▲	●	●	●	▲	●	▲	●	●	●	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Rice	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Washington University in St. Louis	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	▲	●	●	●	▲	●	▲	●	●	●	▲	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Emory	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
University of Notre Dame	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
University of California-Berkeley	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Vanderbilt	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	▲	▲	▲	●	▲	▲	▲	▲
University of Virginia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	▲	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Carnegie Mellon	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	●	●	▼	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	●	▲	▲	▲	▲	●	▲	▲	▲	▲
Georgetown	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
University of California-Los Angeles	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	●	●	▼	●	▼	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	●	▲	▲	▲	▲	●	▲	▲	▲	▲
University of Michigan-Ann Arbor	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	●	●	●	●	▲
University of North Carolina-Chapel Hill	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Wake Forest	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
College of William and Mary	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	●	●	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Tufts	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Brandeis	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
University of California-San Diego	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
University of Rochester	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Case Western	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Lehigh	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
New York University	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●
University of Illinois-Urbana-Champaign	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	▼	▼	●	●	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●
University of Wisconsin-Madison	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●
Boston College	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	▼	▼	▼	●	●	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●
Georgia Institute of Technology	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	▼	▼	▼	●	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●
Pennsylvania State University	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●
University of California-Davis	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●
University of Southern California	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●
Tulane	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●
University of California-Santa Barbara	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●
University of Texas-Austin	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●
University of Washington	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●
Yeshiva	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●
University of California-Irvine	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●
University of Florida	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	▼	▼	●	▼	▼	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	

ranked thirteenth with a score of 86. Only schools ranked below thirteenth all have scores that are significantly lower than Stanford's.

In general, when the overall score for a school is compared to that of every other school in its ranking, two groups emerge: schools that *either* score significantly higher or lower than the school and schools with scores that are not significantly different. For some schools that fall towards the middle of the overall score range, three groups emerge: schools that score significantly higher, schools that score significantly lower, and schools with scores that are not significantly different. This pattern is consistent across all the comparison tables.

For example, among the business schools in Table 29, three distinct groupings emerge. The first group comprises 12 schools at the top of the rankings, extending from first-ranked Stanford to twelfth-ranked Dartmouth College. These schools have scores that are not significantly different from each other but that are significantly higher than all other schools' scores. The second grouping extends from thirteenth-ranked New York University to eighteenth-ranked University of Texas-Austin. These schools have scores that are not significantly different from each other but that are significantly lower than the top-ranked schools in the first group and significantly higher than the lower-ranked schools in the third grouping. The third group is the largest. It comprises 31 schools, extending from twentieth-ranked Purdue University-West Lafayette to forty-ninth-ranked Brigham Young and Pennsylvania State Universities. These schools all have scores that are not significantly different from each other but that are significantly lower than the scores of schools in the first two groups.

There are also three groupings among the schools of education in Table 30, albeit not as clear as those observed among business schools. The first group comprises the top-three-ranked schools of education – Harvard University, Teacher's College, and Stanford University. These schools have scores that are not significantly different from each other but that are significantly higher than the scores for almost all other schools in the top fifty. The second group of schools extends from fifth-ranked University of California-Berkeley to thirteenth-ranked Indiana University-Bloomington, Michigan State University, and the University of Virginia. These schools have scores that are not significantly different from each other but that are significantly lower than the top-three schools and significantly higher than some of the lower-ranked schools. The final group of schools comprises 36 schools – from sixteenth-ranked New York University and the University of Oregon to the six schools tied for forty-sixth rank. These schools all have scores that are not significantly different from each other but that are significantly lower than the scores of most schools in the top two groups.

The top group among the law schools in Table 31 consists of four schools – Yale, Harvard, Stanford and New York Universities – that have statistically similar scores and that score significantly higher than almost all other schools. The second group extends from fifth-ranked Columbia University to twenty-eighth-ranked Emory University, and consists of schools with scores that are not significantly different from each other but that are significantly lower than the top four schools and significantly higher than many lower-ranked schools. The third group consists of 25 schools, ranging from twenty-ninth ranked Brigham Young University to the four schools tied for fiftieth rank. These schools have scores that are not significantly different from each other but which are statistically lower than almost all other schools in the first two groups.

The national liberal arts colleges in Table 32 can also be placed into three groups. The first group comprises the top-eight-ranked schools. These schools have scores that are not significantly different from each other but that are significantly higher than almost all other schools. The second group comprises 13 schools, from ninth-ranked Bowdoin to the four schools tied for eighteenth rank. These schools have statistically similar scores but their scores are significantly lower than the top-ranked group and significantly higher than the 21 schools in the third group. Schools in the third group all have scores that are not significantly different from each other but that are significantly lower than the scores for the 21 schools in the top two groups.

Finally, three similar groupings emerge among the national universities in Table 33. The first grouping extends from first-ranked California Institute of Technology to fourth-ranked Princeton and Yale Universities. These schools all have scores that are not significantly different from each other but that are significantly higher than almost all other schools' scores. The second grouping begins at sixth-ranked Stanford and extends to twenty-fifth-ranked University of Michigan-Ann Arbor. These 21 schools all have scores that are not significantly different from each other but that are significantly lower than the top five schools and significantly higher than the lower-ranked schools in the third grouping. The third grouping comprises 22 schools that all have scores not significantly different from each other but significantly lower than most other schools' scores.

These groupings can be used to interpret a school's rank. For example, while Stanford is ranked first among business schools by *US News* and appears to have a higher overall score than the other 49 schools in the rankings, the results in Table 29 would suggest that Stanford's score is really only higher than 37 of these schools and that it ties for first rank with 12 schools. While Sarah Lawrence is ranked thirty-ninth out of 42 schools in the top 40 national liberal arts colleges, the results in Table 32 would suggest that only 21 schools rank higher than Sarah Lawrence and that the remaining 20 have statistically similar scores that place them at the same rank. Thus, instead of being ranked thirty-ninth, Sarah Lawrence could be ranked twenty-second, along with 20 other schools.

The discrepancy between these results and the *US News* rankings is not as large as it first appears. While the rankings appear to spread schools out in individual ranks, several small groupings of schools emerge in the form of ties. A tie occurs when schools have the same overall score. Since *US News* began rounding overall scores to the nearest whole number in 1998, the number of ties or small groupings in the rankings has increased dramatically.

For example, as Table 34 illustrates, while *US News* presents a list of the top 40 or top 50 schools in each ranking category (i.e., business, education, law, national liberal arts college and national university), there are between nine (national liberal arts) and 16 (law school) ties for rank in these listings, with between two (all listings) and eight (business school) schools tied for a particular rank. This means that the number of schools that actually have their own distinct rank is smaller than it at first appears –

varying between 13 (law school) and 17 (national liberal arts). The ties for rank tend to be towards the middle and lower range of the scores where there are much smaller differences between schools' scores.

Table 34
Ties for Rank and Score Ranges for the Top-fifty Schools in Each Ranking

	Business	Education	Law	National Liberal Arts College	National Universities
Number of Ties for Rank (a)	12	14	16	9	14
Number of Non-tied Ranks (b)	14	15	13	17	14
Total Number of Ranks (a + b)	26	29	29	26	28
Score Range	39	53	48	32	39

Because there tends to be a bigger gap between the overall scores of schools at the top of the rankings and between the overall scores for these schools and lower-ranked schools, the t-tests for these comparisons generally find a statistically significant difference (as long as the standard error estimates for each score are not very large). Because the overall scores for the lower-ranked schools tend to clump together more, the t-tests for these comparisons are generally non-significant (as long as the standard error estimates are not very small). This pattern also corresponds to the pattern in Tables 29 through 33 where bands of significantly higher and lower scores tend to occupy the lower left and upper right sides of the table and a "square" of non-significantly different scores tends to emerge in the lower right-hand corner.

Summary and Conclusion

The results of the factor analyses indicate the present of a unidimensional structure in the indicators used for most of the rankings. The presence of a unidimensional structure does not necessarily mean that the indicators are measuring academic quality. Confirmation of this would require further tests of the relationship between a school's performance on these indicators and other established measures of academic quality. In addition, results suggest the presence of two to three dimensions underlying the overall score for most rankings. These dimensions do not correspond to the categories that *US News* uses to create sub-rankings for schools, suggesting that it might be better to combine some of the current categories into broader conceptual categories before computing sub-rankings in these areas.

The results of the comparability analyses suggest that a variety of changes have taken place in the indicators used to construct the rankings over the last five editions with most of these changes occuring in indicator definition and methodology. In addition, some of the rankings – e.g., law – have experienced more changes than others. There does not appear to be a relationship between the amount of change in indicators and the amount of change in rank experienced by schools during this time period. For example, while not experiencing much change in indicators used, the schools of education have experienced markedly more change in their ranks over time than other schools. It is not evident why this is the case. Access to the complete data set used to create the rankings would be required in order to shed further light on this issue.

The results of the freedom from error analyses suggest that the error surrounding the overall scores used to rank schools is sufficient to create overlap between these scores. The comparison tables show that the degree of overlap is quite large – particularly in the case of schools of education. In addition, these results suggest that some schools are quite robust to changes in indicators, while others are quite influenced by them, and this is represented by smaller standard errors for the former and larger for the latter.

Chapter Five

Discussion and Conclusion

In Chapter One, some of the major concerns about college rankings were summarized by Gerhard Casper, Chancellor at Stanford, who pointed out the "false precision" of the overall scores, the "tinkering to 'perfect' the weights and formulas," and the lack of recognition of the fact "that quality may not be truly quantifiable." It was noted that not much research exists that addresses these concerns and the associated technical issues of freedom from error, comparability, and validity. This study filled that gap by carrying out empirical analyses of the technical issues involved, using data from the last five editions of the *US News* business, education, engineering, law, medicine, primary-care, national liberal arts college, and national university rankings. The results of these analyses will now be discussed in terms of their implications for the debate over the *US News* rankings. The results will also be used to discuss the implications for rankings of educational institutions in general and useful directions for further research.

Unidimensionality

Casper's concern "that quality may not be truly quantifiable" goes to the heart of the rankings debate. Apart from questioning the very concept of college rankings, the main issue raised is whether current rankings are actually measuring academic quality. As an initial step towards answering this question, this study addressed the issue of whether the indicators used are measuring a common attribute – one that may or may not be "academic quality." Results indicate that there is a strong unidimensional structure underlying the indicators used for most of the rankings. In addition, the results show that the rotated factor structures do not correspond to the categories that *US News* uses to organize indicators. Both of these findings are weakened by the fact that indicators were missing for some of the rankings.

Critics of the rankings would not necessarily see these finding as weakening their position since the presence of a unidimensional structure does not prove that this structure is "academic quality." *US News* might see the results as lending credence to their use of an overall score to construct their rankings, although their sub-rankings are not supported. In either case, the findings must be viewed as a first step towards addressing the bigger issue of the overall validity of the rankings.

Comparability

Critics have often taken *US News* to task over their constant "tinkering to 'perfect' the weights and formula" used to create the rankings, citing these changes as making it impossible to interpret a school's change from year to year in the rankings in any substantive fashion.

The results of this study show that the amount of change in indicators has been quite modest for some of the rankings (e.g., the national liberal arts college and national university rankings) but quite large for some others (e.g., the law rankings). The large amount of change in the law rankings is strong support for critics' concerns over the lack of comparability from year to year in the rankings. However, the reason the law rankings experienced more change than others is no surprise to *US News* – it is partly because more new sources of information became available on law schools than for other types of schools and *US News* responded to this availability of improved information by changing the ranking's indicators. While this striving for improvement is understandable, the usefulness of the rankings is weakened by the lack of a core set of indicators that could be used to make comparisons across years while still allowing *US News* to make modifications to the overall formula.

The fact that there has been little change overall in the top 50 ranked schools for most rankings between 1995 and 2000, and that there is no apparent relationship between the amount of change in a ranking's indicators over this period and the amount of change in its top 50 ranking, seems to indicate that schools are generally robust to changes in indicators in terms of their final overall score and placing in the rankings. However, the freedom from error analyses indicate that the precision of this overall score is illusory and that some schools are more affected by changes in indicators than others. In addition, the results of the comparability analyses do not explain why schools of education seem to have experienced more change in their relative academic quality across years than other types of schools. Here, as with many of the other analyses, the question becomes impossible to answer as the data are not available.

Freedom from Error

One of the biggest criticisms of the rankings is the "false precision" implied by the use of an overall score to assign schools to discrete ranks. The results of this study certainly call into question the use of overall scores (albeit rounded to the nearest whole number) to assign school to individual ranks. These analyses show that when you combine the fairly narrow range of scores for the top 50 schools (generally no more than one or two points separate successive schools) with their standard errors (an average standard error of 2.43 for the top 50 business schools, 5.54 for the top 50 schools of education, 1.91 for the top 50 law schools, 2.17 for national liberal arts colleges, and 2.29 for national universities), precision blurs and schools start to group in bands rather than discrete ranks. The results confirm the critics' sense of unease at the precision of a single score, particularly in the case of the education rankings. They also suggest that year-to-year changes in rank may be partly a function of the amount of uncertainty that exists in any attempt to measure.

Implications of the Results

The implications of these results can be discussed in terms of the larger debates over college rankings, educational rankings at the elementary and secondary level, and rankings in general.

In terms of the debate over college rankings, the results suggest that technical analyses are a useful accompaniment to conceptual arguments over rankings. The three analyses provide quantitative information that can be used in discussing concerns with validity, change, and error in the rankings. The analyses also highlight a problem with the current data made available by *US News* – there simply is not enough of it. If the debate over rankings is to be carried out in a constructive manner, more information on the data used to create the rankings will have to be made available. If space constraints mean that not all data can be printed in the rankings publications, the data should be available on request to researchers or put on the *US News* rankings website.

The results of this study are also suggestive in light of the current movement at the elementary and secondary levels to use test scores and other indicators to rank schools and school districts in certain states. The following questions that revolve around the issues of validity, comparability, and freedom from error, should be asked by those attempting such rankings:

- What studies will be carried out to ensure that these indicators are actually measuring what they are supposed to measure and that they do this for all types of schools and school districts?
- How will change be measured? In particular, how will changes in the measures used to rank schools be separated from change in a school's performance across years?
- If schools are to be placed into categories based on their overall score, what allowance will be made for error in the score? (This issue is particularly important in the case of schools bordering two categories.)
- Are the rankings and their associated benchmarks applied similarly across all schools or is a school's own circumstances and mission also taken into account?

This study's results can also be applied to the interpretation of rankings in general. In particular, the results can be summarized in the form of the following guidelines for consumers:

- Indicators used to construct a ranking should be examined in light of whether they actually represent what they are supposed to represent – e.g., quality or value for money.
- The score difference between products that are near or beside each other in the rankings should be examined in order to assess the actual magnitude of the

distance between them. Products that are separated by only a few points might perform very similarly. All else being equal, it might then be better to choose the cheaper product.

- The weights assigned to the various indicators used to compute the rankings should be evaluated in terms of whether they match the consumer's needs and values.

Recommendations for College Ranking Methodology

College rankings are attractive, both as revenue generators and as a tool for students and parents. They are a convenient way to condense and present a variety of information on colleges. Thus, they are probably going to become even more popular in today's increasingly complex and information-driven society. Based on the analyses conducted for this study, as well as a review of the relevant literature on the subject, the following are recommendations for improving the methodology used to create these rankings:

- Include more student-outcome indicators of academic quality: Current academic quality rankings tend to rely on information that is already available and easily quantifiable – e.g., test scores and class size – and that is primarily input-based (i.e., something that students or faculty bring with them to the institution). However, in order to more accurately represent the concept of "academic quality" as the quality of a school's graduates, not merely the quality of incoming students, *US News* and others need to include more short- and long-term student-outcome indicators such as employment rates, civic involvement and entry to graduate school. These are especially needed for the undergraduate rankings. Some ways to obtain more of these types of measures would be to embed them in a general survey of higher education institutions or to require schools to track students four years or more beyond graduation. The information obtained could then be made available to the various organizations involved in ranking efforts.

- Retain a core set of indicators: A vital characteristic of annual rankings should be the ability to make comparisons across years. Otherwise, they become misleading. Therefore, *US News* and others who produce rankings at regular intervals should maintain a core of indicators and present information on schools' change in performance on these indicators across years. However, it should be made clear to the reader that change on these specific indicators does not necessarily represent an overall upward or downward change in a school's academic quality. If *US News* could introduce greater stability into their overall ranking methodology, it would also be possible to make some statements about change in a school's rank from one year to the next.

- Include an error estimate: A college ranking, by its very name, requires that schools be ordered in a line according to some criterion. However, to signal the amount of uncertainty surrounding this process to the reader, and to allow for more accurate interpretations of the information presented, an error estimate should be used to indicate the amount of overlap between the scores of neighboring schools. The degree of uncertainty or possible error around a school's score could be signified by presenting the possible range of scores for each school (using plus and minus two standard errors as the guide) or by grouping schools in bands.

- Include a criterion-referenced component: Rankings are relative in nature. A school's performance on individual indicators, and overall, is evaluated in terms of how it compares to the performance of other schools. This can lead to a situation where a school may perform well on an indicator but still be ranked low on that indicator if other schools perform much better. The same applies to a school's overall score and subsequent rank. To lessen the extent to which such comparisons can lead to inaccurate conclusions about the academic quality of a school, overall or in a particular area, it would be useful to establish benchmarks of quality in certain areas – e.g., what is a poor, fair, good or excellent freshmen retention rate? This exercise is also useful as it requires one to think about the value of the information and its appropriateness as an indicator of academic quality. For example, is it possible to establish what makes for a poor, good or excellent alumni giving rate in a way that makes sense across all institutions in a ranking? If not, why is this indicator being used to compare their academic quality?

Acting on the above recommendations would produce rankings that benefited several constituencies, including media, educational, research, consumer and higher education groups. In particular, implementing the above recommendations could help improve the credibility of the media rankings (thereby boosting sales), help provide useful and more accurate information for consumers for improved decision-making, and provide better information to the higher education community itself to aid in self-evaluation, setting goals for improvement, and monitoring change over time.

Recommendations for Future Research

The analyses presented in this study represent a starting point for exploration of the extent to which the assumptions upon which the rankings are based actually stand up to technical scrutiny. The following are recommendations for future research in this area:

- Replication of the current study: The current study needs to be replicated using the full *US News* dataset. This will illustrate whether access to the full set of indicators used for each ranking will significantly change the results of the factor analyses and the jackknife standard error calculations. In addition, this study should be replicated using data for different years and could be extended to include the regional rankings at the undergraduate level.

- Studies of external validity: In order to address the issue of whether the rankings actually measure the academic quality of colleges, studies on their external validity need to be conducted. These studies could examine the degree of correspondence between the *US News* rankings and other measures of academic quality – e.g., parents' and students' ratings of colleges' academic quality both before and after enrolling at a particular college, surveys of students five and ten years out from college graduation, and other academic quality rankings.

- Momentum studies: It would be useful to explore the extent to which there is a certain momentum attached to a school's rise or fall in the rankings. For example, if a school falls in rank, does it continue to fall in subsequent years? What does it take to halt or reverse such movement? Is improvement on certain indicators key to halting or reversing such a trend?

- Rankings primer: A rankings primer could be produced for consumers of rankings in general or for users of college rankings information in particular. The goal of the primer would be to educate consumers on the basic assumptions underpinning rankings and the kinds of questions they should ask when using them.

Conclusion

This study addressed some technical issues surrounding one ranking in particular – *US News'* rankings of colleges and graduate programs. Technical issues aside, what is seen by some as especially inappropriate about the academic quality rankings produced by *US News* and others is the way that the media have effectively taken over the evaluation of the academic quality of higher education and the fact that their efforts are very appealing to a public that wants quick, accessible information on colleges and graduate schools. If the influence of the media rankings is to be softened or removed, there must be a conscious effort by those in higher education to provide the public with comparable, accurate information on colleges, on a regular basis, and in a format that will appeal to the public. A coordinated effort might require the involvement of a national organization in order to ensure an appropriate level of consistency and objectivity.

Another way to temper the contentious nature of the rankings would be for *US News* to follow the model adopted by *Consumer Reports* with their product rankings. *Consumer Reports* ranks products, but does not allow the product manufacturers to use the rankings in their advertising. Similarly, *US News* could continue to rank colleges and graduate programs but not allow institutions or others to use the rankings in their advertising or other promotion strategies.

In addition, there should be a re-assessment of the meaning of academic quality in contemporary society. This is an important exercise as the hand of history has made academic quality equivalent to the venerable aspect of schools. We are now living in a different era, where quality is tied up with the match between a student and a school. What might another system of ranking information look like that took this more contemporary meaning into account?

Finally, the consumer needs to be educated. Students and their parents need to learn how to pick the college that's right for them and also how to be a critical and analytical consumer of the information on colleges provided by the media, by individual schools, and by society in general. Certainly, the answer to the debate over the rankings does not fall with any one group or action. Rather, it requires the involvement of those who produce the rankings (currently the media), those who are ranked by them (the colleges) and those who use them (the public).

References

Arenson, K. W. (1998). New York seeks to assess colleges and universities. *New York Times*, September 27.

Art and Science Group, (1995). Influence of *US News* rankings on college choice, *StudentPOLL: Market Intelligence for Higher Education*, 1(1).

Astin, A. (1961). A re-examination of college productivity. *Journal of Educational Psychology*, 52(3), June, pp:173-178.

Astin, A. (1962). Productivity of undergraduate institutions. *Science*, 136, April 13, pp:129-135.

Babcock, K. C. (1911). Higher education in the United States. *Report of the Commissioner of Education for the Year Ended June 30, 1911*, vol. 1. Washington, DC: Government Printing Office, pp: 37-65.

Beaton, A. E. (1997). Missing scores in survey research. In J. P. Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook*, pp. 763-766. New York: Elsevier Science.

Beaton, A. E., and Zwick, R. (1992). Overview of the National Assessment of Educational Progress. *Journal of Educational Statistics*, 17(2), pp. 95-109.

Becker, G. S. (1993). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* Chicago: University of Chicago Press.

Blair, J. (2000). Study: College rankings affect aid packages. *Education Week*, 19(19), pp. 6-7.

Blau, P. M., and Margulies, R. Z. (1974-75). The reputations of American professional schools. *Change*, 6(10), December/January, pp: 42-47.

Bodenhorn, H. (1997). Teachers, and scholars too: economic scholarship at elite liberal arts colleges. *The Journal of Economic Education*, 28(Fall 1997), pp. 323-36.

Braxton, J. M., and Nordvall, R. C. (1996). An alternative definition of quality of undergraduate college education: toward usable knowledge for improvement. *The Journal of Higher Education*, 67 (Sept./Oct.), pp: 483-97.

Camilli, G., and Firestone, W. A. (2000). Values and state ratings: An examination of the state-by-state education indicators in quality counts. *Educational Measurement: Issues and Practice*, 18(4), pp. 17-25.

Cantor, G. (1996). Universities increasingly question the criteria of college rankings. *The Detroit News*, December 1.

Caples, A. (1999). Law school rankings out: We're still number two: US News survey a "beauty pageant" says Chmura. *Harvard Law Record*, April 9, pp. 1, 7.

Carmody, D. (1987). Picking College Guides: No Easy Task. *New York Times*, December 2, 1987, v.137, p.19

Cartter, A. (1966). *An Assessment of Quality in Graduate Education*. Washington, DC: American Council on Education.

Cassidy, T. (1999). Schools will be ranked on MCAS scores. *Boston Globe*, August 9, p. A01.

Cattell, J. McKeen (1906). *American Men of Science: A Biographical Dictionary*. New York: Science Press.

Cattell, J. McKeen (1910a). *American Men of Science*, 2nd ed. New York: Science Press.

Cattell, J. McKeen (1910b). A statistical study of American men of science. J. McKeen Cattell (ed.), *American Men of Science*, 2nd ed. New York: Science Press, pp: 537-563.

Cattell, J. McKeen (1910c). A further statistical study of American men of science. J. McKeen Cattell (ed.), *American Men of Science*, 2nd ed. New York: Science Press, pp:564-596.

Conrad, C., and Webster, D. S. (1986). Using faculty research performance for academic quality rankings. *New Directions for Institutional Research*, 50, pp: 43-57.

Cornwell, T. (1998). Law league table debunked. *The Times Higher Education Supplement*, no. 1321, February 27, p. 60.

Crenshaw, A. B. (1999). Colleges by the numbers. *Washington Post*, August 29, p. H01.

Daniel, L. G. (1990, November). *Common Factor Analysis or Components Analysis: An Update on an Old Debate*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED325 531)

Daniel, L. G. (1991). Operationalization of a frame of reference for studying organizational culture in middle schools. In B. Thompson (Ed.), *Advances in*

Educational Research: Substantive Findings, Methodological Developments (pp. 1-24). Greenwich, CT: JAI.

Doctoral Faculty Decade Publication Project (1999). Doctoral faculty decade publication project. *Journal of Education for Social Work.*

Editorial (1997). America's 'best' colleges? *The Christian Science Monitor*, January 7, p. 20.

Efron, B., and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* New York: Chapman and Hall.

Evaluation News, 2(1), February 1981, pp. 85-90.

Finucane, M. (1999). AG Reilly: Do school grading the right way. *Boston Globe*, September 9, p.

Flexner, A. (1910). *Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching*, Bulletin no. 4. New York: Carnegie Foundation for the Advancement of Teaching.

Geraghty, M. (1997). US News alters methodology of its controversial college rankings. *The Chronicle of Higher Education*, 43 (July 18), p. A34.

Gies, W. J. (1926). Dental Education in the United States and Canada: A Report to the Carngie Foundation for the Advancement of Teaching, Bulletin no. 19. New York: Carnegie Foundation for the Advancement of Teaching.

Gilbert, J. (1992). Satisfying a hunger for information: A response to B. Ann Wright. *The College Board Review*, No. 163, pp. 16-17, 35-36.

Glass, G. V., and Hopkins, K. D. (1996). *Statistical Methods in Education and Psychology*, 3rd edition. Needham Heights, MA: Simon and Shuster.

Gledhill, L. (1999). Davis gets ok on plan to rank schools: Index called key to improvement goal, San Francisco Chronicle, November 10, p. A3.

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

Gorsuch, R. L. (1974). *Factor Analysis.* Philadelphia, PA: W. B. Saunders.

Gray, H. L. and Schucany, W. R. (1972). *The Generalized Jackknife Statistics.* New York: Marcel Dekker.

Guskey, T. R., and Kifer, E. W. (1989). Ranking school districts on the basis of statewide test results: Is it meaningful or misleading? Paper presented at the *Annual Meeting of the American Educational Research Association*, San Francisco, CA, March.

Hair, J. F., Anderson, R. E., Tatham, R. L., and Black, W. C. (1998). *Multivariate Data Analysis*, 5th edition. Upper Saddle River, NJ: Prentice Hall.

Hattendorf, L. C. (ed.) (1993). *Educational Rankings Annual.* Detroit, MI: Gale Research Inc.

Hattendorf, L. C. (1996). Educational rankings of higher education: Fact or fiction? Paper presented at the *International Conference on Assessing Quality in Higher Education*. Queensland, Australia, July 15.

Hicks, N. (1997). Report: Creighton still No. 1 in Midwest. UNL moves up this year in US News and World Report's annual ranking of colleges. *Omaha World-Herald*, August 22, p. 13sf.

Hinkley, D. V. and Wei, B. C. (1984). Improvement of jackknife confidence limit methods. *Biometrika* 71, pp. 331-339.

Hirsch, T. M., and Miller, T. R. (1991). *Investigation of the Dimensional Structure of the P-ACT+*. American College Testing Research Report Series. Report No. ACT-RR-91-1. Iowa: American College Testing Program.

Jensen, E. L., and Webster, D. (1981). Studies of graduate sociology departments in "The American Sociologist": 1965-1980. Paper presented at the *Annual Meeting of the American Educational Research Association*. Los Angeles, CA. April 13-17.

Johnson, Johnson, Gott, and Zimmerman (1997). Assessing the validity of scores on the Charles F. Kettering scale for the junior high school. *Educational and Psychological Measurement*, 57(5), pp. 858-869.

Johnson, W. R. and Neal, D. (1998). Basic skills and the black-white earnings gap. In C. Jencks and M. Phillips (eds.), *The Black-White Test Score Gap*. Washington, DC: Brookings Institution.

Joint Committee on Standards for Educational Evaluation (1994). *The Program Evaluation Standards*, 2nd edition. Thousand Oaks, CA: Sage.

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

Kaiser, H. F. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement,* 34, pp. 111-117.

Katchadourian, H. and Boli, J. (1994). *Cream of the Crop: The Impact of Elite Education in the Decade after College.* New York: Basic Books.

Kim, J., and Mueller, C. W. (1978). *Factor Analysis: Statistical Methods and Practical Issues.* Newbury Park, CA: Sage.

Lawley, D. N., and Maxwell, A. E. (1971). *Factor Analysis as a Statistical Method.* London: Butterworth and Co.

Lay, L. (1996). Colleges question rankings list. *The Baylor Lariat,* December 3.

Levin, S. (1999). Ignore college ranking – become an educated consumer. *Washington Parent Magazine.* Available at: http://www.washingtonparent.com/articles/9712/rankings.htm

Machung, A. (1998). What's at stake: College rankings and the new media. *Change*, July/August, pp. 13-16.

Maclean, A. H. (1900). *Where We Get Our Best Men.* London: Simpkin, Marshall, Hamilton, Kent, and Co.

McCall, A. (September, 1999). Cambridge is still in a class of its own. *The Sunday Times University Guide*, 19 September, pp. 4-6.

McDonough, P. M., Antonio, A. L., Walpole, M., and Perez, L. X. (1998). College rankings: Democratized knowledge for whom? *Research in Higher Education*, 39(5), pp. 513-37, October.

Miller, R. G. (1964). A trustworthy jackknife. *Ann. Math. Statist.* 39, pp. 1594-1605.

Miller, R. G. (1974). The jackknife – a review. *Biometrika* 61, pp. 1-17.

Monks, J., and Ehrenberg, R. G. (1999). *The Impact of US News and World Report College Rankings on Admissions Outcomes and Pricing Policies at Selective Private Institutions.* Cambridge, MA: National Bureau of Economic Research.

Mufson, S. (1999). Rankings all-important to GWU. *Washington Post*, Sunday, March 14, p. A01.

Mulaik, S. A. (Ed.) (1992). [Theme issue on principal components analysis]. *Multivariate Behavioral Research*, 27(3).

Norusis, M. J. (1990a). *SPSS Base System User's Guide*. Chicago, IL: SPSS Inc.

Norusis, M. J. (1990b). *SPSS Introductory Statistics Student Guide*. Chicago, IL: SPSS Inc.

Nunnally, J. (1978). *Psychometric Theory*. New York: McGraw-Hill.

Oslin, R. (1999). BC again in US News' Top 40. *The Boston College Chronicle*, 8(1), September 3, p.6.

Pellegrini, F. (1999). Those bouncing college rankings – a 101. *Time*, August 31.

Poffenberger, A. T. (1947). *James McKeen Cattell: Man of Science, 1980-1944*. Lancaster, PA: Science Press.

Powers, S., and Gallas, E. J. (April, 1980). *The Factorial Structure and Item Composition of the Comprehensive Tests of Basic Skills, Reading Comprehension Test, When Administered to ESEA Title 1 Sixth and Seventh Grade Students.* Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Quenouille, M. (1949). Approximate tests of correlation in time series. *J. Royal. Statist. Soc. B*, 11, pp. 18-44.

Richtel, M. (1998). University takes magazine to task on the internet. *New York Times*, August 27, Section G, p. 3.

Robinson, G. E., and Brandon, D. P. (1994). *NAEP Test Scores: Should they be Used to Comapre and Rank State Educational Quality? ERS Concerns in Education.* Arlington, VA: Educational Research Service.

Rogers, E. and Rogers, S. (1997).

Roose, K. D., and Andersen, C. J. (1970). *A Rating of Graduate Programs.* Washington, DC: American Council on Education.

Salant, P., and Dillman, D. A. (1994). *How to Conduct Your Own Survey.* New York: John Wiley and Sons, Inc.

Sanoff, A. (1998). Rankings are here to stay: colleges can improve them. *The Chronicle of Higher Education*, 45(2), September 4, p. A96.

Scriven, M. (1991). *Evaluation Thesaurus*, 4th edition. Newbury Park, CA: Sage.

Seaman, B. (October 26, 1998). What makes a good college. *Time*, 152(17), pp:1,2

Shea, C. (1995). Rankings that roil also gain influence. The Chronicle of Higher Education, 42 (Sept 22), pp: A53, A54.

Smetanka, M. J. (1998). Magazine's new ratings on colleges don't rank high at "U". *Star Tribune*, February 20, p. 7B.

Stecklow, S. (1995). Cheat sheets: Colleges inflate SATs and graduation rates in popular guidebooks: Schools say they must fib to US News and others to compete effectively: Moody's requires the truth. *The Wall Street Journal*, Wednesday, April 5, pp. 1, A8.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, 52, pp. 589-617.

Thompson, B., and Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. Educational and Psychological Measurement, 56(2), pp. 197-208.

Thompson, J. J., and Morse, R. J. (1998). An explanation of the *US News* rankings: Putting the numbers into context. *US News and World Report America's Best Colleges 1998*, pp. 66-68.

Thompson, N. (1996). Colleges flunk the trendy ratings game; Education: The US News rankings is subjective, yet schools grovel for a good position and prospective students are misled. *Los Angeles Times*, October 25, Metro, Part B, p. 9.

Toch, T. (1999). Outstanding high schools. *US News and World Report*, 126(2), January 18, pp.46-51.

Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

Tukey, J. W. (1958). Bias and confidence in not quite large samples. (Abstract.) *Ann. Math. Statist.*, 29, pp. 614.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

United States Bureau of Education (1917). *Negro Education: A Study of the Private and Higher Schools for Colored People in the United States*, 2 vols, Bulletin 1916, nos. 38 and 39. Washington, DC: Government Printing Office.

US News and World Report (1983-2000). *America's Best Colleges.*

US News and World Report (1987-2000). *America's Best Graduate Schools.*

Webster, D. S. (1981). Early psychological studies of "genius" and academic quality rankings. Paper presented at the *Annual Meeting of the American Psychological Association*. Los Angeles, CA. August 25.

Webster, D. S. (1986a). *Academic Quality Rankings of American Colleges and Universities.* Springfield, IL: Charles C. Thomas.

Webster, D. S. (1986b). Ranking academic quality: The undergraduate story. *Change*, 18(6), pp. 34-41.

Webster, D. S. (1990). Continuity and change in the highest-ranked PhD-granting departments of English, 1925-1982. Paper presented at the *Annual Meeting of the American Educational Research Association*, Boston, MA. April 16-20.

Webster, D. S. (1992). Rankings of undergraduate education in US News and Money – Are they any good? *Change*, 24(2), pp. 18-31.

West, C. K. (1978). Productivity ratings of institutions based on publication in the Journals of the American Educational Research Association: 1970-1976, *Educational Researcher*, 7(2), pp. 13-14.

Wright, B. A. (1990-91). The rating game: How the media affect college admission. *The College Board Review*, No. 158, Winter, pp. 12-17, 31.

Wright, B. A. (1992). A little learning is a dangerous thing: A look at two popular college rankings. *The College Board Review*, No. 163, Spring, pp. 6-16.

Wu, C. F. J. (1989). Jackknife, bootstrap and other resampling plans in regression analysis (with discussion.) *Ann. Statist.* 14, pp. 1261-1350.

Appendix A

US News Indicators for the 2000 Undergraduate Rankings

This appendix contains actual text from the *US News* college rankings website describing the seven categories of indicators used for the *US News* Undergraduate Rankings and providing a rationale for why each indicator was chosen. Further information can be found at:

http://www.usnews.com/usnews/edu/college/rankings/collmeth.htm

Academic Reputation

The *US News* ranking formula gives greatest weight (25 percent) to reputation because a diploma from a distinguished college so clearly helps graduates get good jobs or gain admission to top graduate programs. Much as an interview allows admissions officials to gauge an applicant's personal qualities, the reputation survey allows respondents to account for a variety of intangibles such as faculty dedication to teaching or student tolerance of cheating.

A school's reputation is determined by surveying the presidents, provosts, and deans of admission at institutions in the same category. Each individual was asked to rate peer schools' academic programs on a scale from 1 (marginal) to 5 (distinguished). Those who didn't know enough about a school to evaluate it fairly were asked to mark "don't know." Market Facts Inc., an opinion-research firm based near Chicago, collected the reputational data; 68 percent of the 3,966 people sent questionnaires responded.

Retention

The higher the proportion of freshmen who return to campus and complete their studies, the better a school is likely to be at offering the classes and services students need to succeed. This measure has two components: six-year graduation rate (80 percent of the retention score) and freshman retention rate (20 percent). The graduation rate indicates the average proportion of a graduating class that earns a degree in six years or less; we considered freshman classes that started between 1989 and 1992. Freshman retention indicates the average proportion of freshmen entering between 1994 and 1997 who returned the following fall. While critics charge that these measures favor schools that "dumb down" their curriculum, students fail to return or graduate for many reasons. Policies to nurture freshmen and support struggling students can make a big difference.

Faculty Resources

Research shows that the more satisfied students are with their contact with teachers, the more they will learn and the more likely it is they will graduate and attend graduate and professional programs.

We use five factors from the 1998-99 academic year to assess a school's commitment to superb instruction. Class size, which accounts for 40 percent of the faculty-resources score, represents the proportion of classes with fewer than 20 students and of classes with more than 50. Faculty salary (35 percent) is the average faculty pay, plus benefits, during the 1997-98 and 1998-99 academic years, adjusted for regional differences in the cost of living (using indices from Runzheimer International). We also weigh the proportion of professors with the highest degree in their field (15 percent); the student-faculty ratio (5 percent); and the proportion of the faculty that is full time (5 percent).

Student Selectivity

The academic climate of a school is determined in part by the abilities and ambitions of the student body. We therefore factor in test scores of enrollees on the SAT or ACT tests (40 percent); the proportion of enrolled freshmen who graduated in the top 10 percent of their high school classes for the national institutions and the top 25 percent for the regional schools (35 percent); the acceptance rate, or the ratio of students admitted to applicants (15 percent); and the yield, or the ratio of students who enroll to those admitted (10 percent of this ranking factor). The data are for the fall 1998 entering class.

Critics charge that some colleges have changed admissions policies to improve their numbers in ways that are harmful to students. For instance, some contend that colleges have emphasized early admission, which requires students to attend a school if accepted, because schools want to increase their yield. As a result, students may abandon their college searches too soon and commit to schools that don't have much incentive to offer generous financial aid. There may be sound economic reasons for schools to push early admission: The policies help schools fill their classrooms and dorms. But making such a move simply to move up a rank or two would be silly; yield accounts for a mere 1.5 percent of the overall ranking.

Financial Resources

Generous per-student spending indicates that a college is able to offer a wide variety of programs and services. *US News* measures the average spending per student on instruction, research, and education-related services during the 1997 and 1998 fiscal years.

Graduation Rate Performance

This indicator of "added value" was developed to capture the effect of the college's programs and policies on the graduation rate of students after controlling for spending and student aptitude. We measure the difference between a school's six-year graduation rate for the class that entered in 1992 and the predicted rate for the class. The predicted rate takes into account the standardized test scores of these students as incoming freshmen and the school's expenditures on them. If the actual graduation rate is higher than the predicted rate, the college is enhancing achievement.

Alumni Giving Rate

The percentage of alumni who gave to their school during the 1997 and 1998 academic years is an indicator of alumni satisfaction.

To arrive at a school's rank, we first calculated the weighted sum of its scores. The final scores were rescaled: The top school was assigned a value of 100, and the other schools' weighted scores were calculated as a proportion of that top score. Final scores for each ranked school were rounded to the nearest whole number and ranked in descending order.

Appendix B

US News Indicators for the 2000 Graduate Rankings

This appendix contains actual text from the *US News* graduate rankings website describing the types of indicators used for the *US News* Graduate Rankings and providing a rationale for why each indicator was chosen. Further information can be found at: http://www.usnews.com/usnews/edu/beyond/gradrank/gbrank.htm

Inputs and Outputs

The statistical measures that account for the greatest proportion of each school's ranking also are indicators that researchers have associated with educational excellence. These measures fall into two categories: inputs, or the qualities that students and schools bring to the educational experience, and outputs, measures of how successfully the program prepared students for success. The data are, with rare exception, collected from the schools themselves.

The caliber of student that a program attracts–which influences the academic climate–is measured by the average undergraduate grade point average and standardized test scores of the entering class. For education and engineering, the Graduate Record Examination is used; for medicine, the Medical College Admission Test; for business, the Graduate Management Admission Test; and for law, the Law School Admission Test.

Other input measures reflect outlays that affect the quality of education. For example, engineering schools report their total research expenditures as well as the number of faculty members engaged in research. Law schools are asked how much they spend per student on faculty salaries, libraries, and student support services. (This year, we adjusted the law expenditures for regional differences in the cost of living.)

To judge how capably a program develops its students, *US News* factors in as many output measures as possible: the rates at which law school graduates pass the bar, for example, and the median starting pay package enjoyed by new M.B.A.'s, including base salaries, guaranteed bonuses, and other forms of compensation. In collecting information, *US News* uses standardized data whenever possible. For example, two years ago, a coalition of business school administrators (the MBA Career Services Council) began asking schools about job placement and starting salaries. *US News* incorporated the council's most recent versions of these questions into the surveys sent to the business schools this year. Similarly, we used language from the American Bar Association's annual survey of law schools when we asked the schools for their enrollment and bar passage rates. This approach increases the likelihood that the rankings are based on accurate information and that all schools report in a consistent manner.

To arrive at a school's rank, we first computed the weighted sum of the school's scores on each quality indicator. The weights reflect *US News's* judgments about which measures of quality matter most. This year, scores for each indicator were standardized before applying the weights. (This accepted statistical adjustment, which recognizes that some indicators vary more around their average value than others, ensures that the formula weights are applied without distortion.) The final scores were rescaled; the top school was assigned a value of 100, and the other schools' weighted scores were calculated as a proportion of that top score. The scores were then rounded to the nearest tenth and ranked in descending order. Detailed information about the quality indicators and weights used to rank schools in each discipline appears on the following pages: business, 28; law, 36; medicine, 51; education, 66; and engineering, 74.

Specialties

Beyond identifying the best overall programs in a discipline, *US News* ranks the top schools in some of the larger specialties, based on the responses of academic experts. Someone who hopes to practice environmental law, for example, can view the top-ranked programs in that specialty. Biomedical engineering? Click here to see the universities with the best programs. In all cases, the schools verified their specialty offerings in writing or, in a few cases, by phone.

Master's and doctoral programs in all other disciplines covered in the guide–the arts, sciences, social sciences, library science, and the allied health fields–are ranked only by reputation. Again, *US News* surveyed the dean or top administrator and at least one other administrator or faculty member at each school and asked them to rate the programs they were familiar with in their particular discipline. The disciplines ranked by reputation only are generally evaluated every third year; the programs assessed this year include doctoral programs in biology, chemistry, computer science, geology, mathematics, and physics, and master's programs in library science. Data for the rankings in the arts were gathered in 1996; for health, in 1996 and 1997; and for the humanities and social sciences, in 1997.

Appendix C

Factor Structures (Pattern Matrices) Following Principal Components Extraction and Oblique (Oblimin) Rotation

Business Indicators (Loading)	Factor Name	Law Indicators (Loading)	Factor Name
Median Compensation (.85) Average GMAT score (.85) Reputation Rank, Academics (-.85) Reputation Rank, Recruiters (-.83) Average Undergraduate GPA (.79) Acceptance Rate (-.76)	Prestige	Reputation Score, Lawyers (.91) Reputation Score, Academics (.90) LSAT (25th percentile) Score (.88) GPA (25th percentile) (.83) Percent Employed at Graduation (.64) Acceptance Rate (.58) Bar Passage Rate (.58) Percent Employed Nine Months After Graduation (.43) Student-faculty Ratio (.12)	General Factor (Not rotated)
Percent Employed at Graduation (.89) Percent Employed Three Months After Graduation (.86)	Employment Success		

Education Indicators	Factor Name	Engineering Indicators	Factor Name
Average GRE, Analytic (.86) Average GRE, Verbal (.83) Average GRE, Quantitative (.76) Acceptance Rate (-.62)	Student Selectivity	Reputation Rank, Engineers (-.91) Number of PhDs Granted (.91) Research Expenditure (.82) Reputation Rank, Academics (-.80)	Reputation and Productivity
Number of Master's Degrees Granted (.84) Number of PhDs Granted (.80) Research Expenditure (.75)	Productivity	Average GRE, Analytic (.84) Average GRE, Quantitative (.80) Acceptance Rate (-.73)	Student Selectivity
Reputation Rank, Superintendents (.93) Reputation Rank, Academics (.88) PhD Student-Faculty Ratio (-.53)	Reputation	Research Expenditure per Faculty Member (.86) Membership in the National Academy of Engineering (.67) PhD Student-Faculty Ratio (.67)	Faculty Profile

Medicine Indicators	Factor Name	Primary-care Indicators	Factor Name
Reputation Rank, Academics (-.90) Research Grants (.86) Average MCAT Scores (.85) Reputation Rank, Directors of Residency Programs (-.80) Average GPA (.78)	Prestige	Average MCAT Scores (-.88) Average GPA (-.84) Student-Faculty Ratio (-.84) Percent of Graduates Entering Primary-care Residencies (.79)	Student Selectivity
Acceptance Rate (.97) Student-Faculty Ratio (-.52)	Capacity	Reputation Rank, Academics (.84) Reputation Rank, Directors of Residency Programs (.74) Acceptance Rate (-.66)	Reputation

1999 National Liberal Arts College Indicators	Factor Name	2000 National Liberal Arts College Indicators	Factor Name
Academic Reputation (.94) SAT (25th percentile) Scores (.93) Top Ten Percent of High School Students (.86) Acceptance Rate (-.82) Freshmen Retention Rate (.77) Graduation Rate (.77) Alumni Giving Rate (.72) Percent of Full-time Faculty (.55)	Student Inputs and Outputs	Academic Reputation (.94) SAT (25th percentile) Scores (.93) Top Ten Percent of High School Students (.89) Acceptance Rate (-.82) Freshmen Retention Rate (.78) Graduation Rate (.77) Alumni Giving Rate (.75) Percent of Full-time Faculty (.58) Percent of Classes with Over 50 Students (.44)	Student Inputs and Outputs
Graduation Rate Performance (-.92) Percent of Classes with Under 20 Students (.77)	Quality of Educational Experience	Graduation Rate Performance (-.85) Percent of Classes with Under 20 Students (-.84)	Quality of Educational Experience
Percent of Classes with Over 50 Students (-.87)	Quality of Educational Experience (2)		

1999 National University Indicators	Factor Name	2000 National University Indicators	Factor Name
SAT (25th percentile) Scores (.95) Top Ten Percent of High School Students (.90) Academic Reputation (.87) Acceptance Rate (-.85) Graduation Rate (.79) Freshmen Retention Rate (.79) Alumni Giving Rate (.77)	Student Inputs and Outputs	SAT (25th percentile) Scores (.96) Top Ten Percent of High School Students (.90) Academic Reputation (.88) Graduation Rate (.83) Freshmen Retention Rate (.83) Acceptance Rate (-.82) Alumni Giving Rate (.80)	Student Inputs and Outputs
Percent of Classes with Over 50 Students (.88) Percent of Classes with Under 20 Students (-.74) Percent of Full-time Faculty (.71)	Quality of Educational Experience	Percent of Classes with Over 50 Students (.88) Percent of Full-time Faculty (.74) Percent of Classes with Under 20 Students (-.72)	Quality of Educational Experience
Graduation Rate Performance (.97)	Value	Graduation Rate Performance (.97)	Value

Appendix D

Changes in Rankings' Indicators over the Last Five Editions of the *US News* Graduate and Undergraduate Guidebooks

Law Schools

Years	Type	Specific Change
1995-1996	Indicator Weight Change	Increased weight given to *Proportion Employed Full and Part Time in Legal and Nonlegal Jobs Six Months After Graduation, Including Those Continuing as Full-time Students and One Fourth of Those Whose Status Was Unknown*
	Indicator Removed	*Ratio of Last Year's On-Campus Job Interview Appointments to the Number of 1994 Graduates*
	Indicator Definition or Methodology Change	*Proportion of the 1994 Graduating Class Employed at Graduation* changed to *Proportion of the 1995 Graduating Class Employed Full and Part Time in Legal and Nonlegal Jobs at Graduation, Including Those Continuing as Full-time Students*
	Indicator Definition or Methodology Change	*Proportion Employed Six Months After Graduation* changed to *Proportion Employed Full and Part Time in Legal and Nonlegal Jobs Six Months After Graduation, Including Those Continuing as Full-time Students and One Fourth of Those Whose Status Was Unknown*
1996-1997	Indicator Definition or Methodology Change	*Proportion Employed Full and Part Time in Legal and Nonlegal Jobs Six Months After Graduation, Including Those Continuing as Full-Time Students and One Fourth of Those Whose Status Was Unknown* changed to *The Proportion of the Class of 1995 Employed Full and Part Time as of February 15, 1996, Including all Those Pursuing Graduate Degrees; and One Fourth of Those Whose Status Was Unknown*
	Indicator Removed	*The Median Starting Salary for 1995 Graduates Employed Only in the Private Sector*
	Indicator Added	*The School's Bar Passage Rate in the Jurisdication Where the Largest Number of Its 1995 Graduates Took the Test for the First Time, Divided by the Overall Passage Rate for First-Time Test Takers in That Jurisdiction*
	Indicator Definition or Methodology Change	*US News* began to compute the *Student-Faculty Ratio* indicator in accord with new American Bar Association guidelines, which credit schools for part-time faculty members and for administrators who also teach
1997-1998	Indicator Definition or Methodology Change	*The School's Bar Passage Rate in the Jurisdication Where the Largest Number of Its 1995 Graduates Took the Test for the First Time, Divided by the Overall Passage Rate for First-Time Test Takers in That Jurisdiction* changed to *The School's Overall Bar Passage Rate for the Summer 1996 and Winter 1997 Tests in the Jurisdiction Where the Largest Number of its 1996 Graduates Took the Test for the First Time, Divided by the Overall Passage Rate for First-Time Test Takers in That Jurisdiction in the Same Time Period*
	Indicator Definition or Methodology Change	*Total Expenditures Per Student for Instruction, Library, and Supporting Student Services During the Year Beginning in Fall 1995* changed to *Average Total Expenditure per Student for Instruction, Library, and Supporting Student Services in the Years Beginning in Fall 1995 and 1996*

	Indicator Definition or Methodology Change	*Financial Aid, Indirect Expenditures, and Overhead per Student* changed to *Average Financial Aid, Indirect Expenditures, and Overhead Expenditures per Student in the Years beginning in Fall 1995 and 1996*
	Indicator Definition or Methodology Change	Change in method used to obtain a school's reputation score by academics
	Indicator Definition or Methodology Change	Change in method used to obtain a school's reputation score by lawyers/judges
1998-2000	Indicator Definition or Methodology Change	The measure of instructional expenditures was adjusted to account for regional differences in cost of living

Business Schools

Years	Type	Specific Change
1995-1996	None	None
1996-1997	Indicator Weight Change	Decrease in weight given to the corporate recruiters' *Reputational Survey*
	Indicator Weight Change	Increase in weight given to the business academics' *Reputational Survey*
	Indicator Definition or Methodology Change	*The Proportion Employed Three Months After Graduation* changed to *The Proportion Employed Three Months After Graduation, Including Students Continuing on for Other Graduate Degrees and One Quarter of Those Whose Status Was Not Known But Excluding MBA Graduates Not Seeking Employment*
1997-1998	Indicator Definition or Methodology Change	Change in method used to obtain a school's reputation score by academics
1998-2000	Indicator Removed	*The Ratio of the Number of Firms that Recruited MBAs on campus to the number of 1997 Graduates*
	Indicator Definition or Methodology Change	*Median Starting Base Salaries, Excluding Sign-Up or Other Performance Bonuses* changed to *Median Total Compensation, Including Starting Base Salaries and Signing Bonuses*
	Indicator Weight Change	Increased weight given to *Employment Rates Three Months After Graduation*

Schools of Education

Years	Type	Specific Change
1995-1996	Indicator Definition or Methodology Change	Inclusion of American Education Research Association fellowships in the *Percent of Faculty Given Any of these Awards* indicator
1996-1997	Indicator Weight Change	Increase in weight given to the academics' reputational survey
	Indicator Weight Change	Decrease in weight given to the superintendents' reputational survey
1997-1998	Indicator Definition or Methodology Change	Change in method used to obtain a school's reputation score by academics
	Indicator Definition or Methodology Change	*The 1996 Dollar Total of Publicly and Privately Funded Research Expenditures, Including Competitive Training Grants, Administered by the Education School* changed to *The Average 1996 and 1997 Dollar Total of the Publicly and Privately Funded Research Expenditures, Including Competitive Training Grants, Administered by the School*
	Indicator Definition or Methodology Change	*Total Research Dollars Divided by the Number of Faculty Members Engaged in Research* changed to *The 1996 and 1997 Average Dollar Amount of Research Conducted by Each Faculty Member at the School In Each of Those Years*
1998-2000	None	None

Schools of Engineering

Years	Type	Specific Change
1995-1996	Indicator Removed	*The Proportion of Engineering Graduate Students Who Were PhD Candidates*
	Indicator Added	*The Proportion of Full-time Engineering Faculty Who are Members of the National Academy of Engineering*
	Indicator Weight Change	Decreased weight given to the *Number of PhDs Granted*
1996-1997	Indicator Weight Change	Increased weight given to the academics' reputational survey
	Indicator Weight Change	Decreased weight given to the practicing engineers' reputational survey
1997-1998	Indicator Definition or Methodology Change	Change in method used to obtain a school's reputation score by academics
	Indicator Definition or Methodology Change	*The 1996 Dollar Total of the Publicly and Privately Funded Research Expenditures Administered by the Engineering School* changed to *The Average 1996 and 1997 Dollar Totals of Publicly and Privately Funded Research Expenditures Administered by the Engineering School*
	Indicator Definition or Methodology Change	*Research Dollar Total Divided by the Number of Faculty Members Engaged in Research* changed to *The 1996 and 1997 Average Dollar Amount of Research Conducted by Each Faculty Member at the School in Each of Those Years*
1998-2000	None	None

Schools of Medicine

Years	Type	Specific Change
1995-1996	None	None
1996-1997	None	None
1997-1998	Indicator Definition or Methodology Change	Change in method used to obtain a school's reputation score by academics
	Indicator Definition or Methodology Change	Change in the number of schools ranked by directors of intern-residency programs
	Indicator Definition or Methodology Change	*Total Dollar Amount of National Institutes of Health Research Grants Awarded to the Medical School and its Affiliated Hospitals in 1996* changed to *The Average Total Dollar Amount of National Institutes of Health Research Grants Awarded to the Medical School and its Affiliated Hospitals in 1996 and 1997*
1998-2000	Indicator Definition or Methodology Change	Change in the number of schools ranked by directors of intern-residency programs

Primary-care Schools

Years	Type	Specific Change
1995-1996	None	None
1996-1997	Indicator Weight Change	Increased weight given to the *Academics' Reputational Survey*
	Indicator Weight Change	Decreased weight given to the *Intern Directors' Reputational Survey*
1997-1998	Indicator Definition or Methodology Change	Change in method used to obtain a school's reputation score by academics
	Indicator Definition or Methodology Change	Change in the number of schools ranked by directors of intern-residency programs
1999-2000	Indicator Definition or Methodology Change	Change in the number of schools ranked by directors of intern-residency programs

National University and National Liberal Arts College

Years	Type	Specific Changes
1996-1997	Indicator Weight Change	Decrease in weight given to the *Retention* category of indicators
	Indicator Added	*Value Added* – a predicted graduation rate was estimated for each school based on test scores of its 1989 entering class and the school's educational expenditures and then compared with the actual six-year graduation rate of the same class
1997-1998	Indicator Definition or Methodology Change	*Total Fiscal 1995 Expenditures for Education Programs Divided by Total Full-time Equivalent Enrollment* changed to *Average of 1995 and 1996 Fiscal Expenditures for Education Programs Divided by Full-time Equivalent Enrollment*
	Indicator Definition or Methodology Change	*Other (non-education) Fiscal 1995 Spending per Student* changed to *Other (non-education) Spending per Student Averaged over 1995 and 1996*
1998-1999	Indicator Definition or Methodology Change	Change in method used to produce the *Academic Reputation* score
	Indicator Removed	*Other (non-education) Spending per Student* is no longer a separate indicator in the Financial Resources category. Aspects of this indicator are combined into the *Average Spending per Student on Instruction, Research, Student Services, and Related Educational Expenditures (Including Libraries) During the 1996 and 1997 Fiscal Years* indicator
	Indicator Weight Change	Full weight for the Financial Resources category given to the *Average Spending per Student on Instruction, Research, Student Services, and Related Educational Expenditures (Including Libraries) During the 1996 and 1997 Fiscal Years* indicator
1999-2000	None	None

Appendix E

Unchanged Indicators Across Rankings, 1995-2000

Ranking	Category	Indicators
Business	**Student Selectivity**	Average Graduate Management Admission Test Scores Average Undergraduate Grade Point Average Proportion of Applicants Accepted
	Placement Success	Employment Rates at Graduation
Education	**Student Selectivity**	Average Graduate Record Examination (GRE) Verbal Scores Average GRE Analytical Scores Average GRE Quantitative Scores Proportion of Applicants Accepted
	Faculty Resources	Current Ratio of Full-time Doctoral Degree Candidates to Full-time Faculty Current Ratio of Full-time Master's Degree Candidates to Full-time Faculty Number of Doctoral Degrees Granted Number of Master's Degrees Granted Proportion of Graduate Students who were Doctoral Candidates
Engineering	**Student Selectivity**	Average GRE Quantitative Scores Average GRE Analytic Scores Proportion of Applicants Accepted
	Faculty Resources	Proportion of Full-time Faculty During the Current Academic Year who Held PhDs Current Ratio of Doctoral Degree Candidates to Full-time Faculty Current Ratio of Full-time Master's Degree Candidates to Full-time Faculty
Law	**Student Selectivity**	Median Law School Admission Test Scores Median Undergraduate Grade Point Average Proportion of Applicants Accepted
	Faculty Resources	Total Number of Volumes and Titles in the Law Library
Medicine	**Student Selectivity**	Average Medical College Admission Test Scores Undergraduate Grade Point Average Proportion of Applicants Accepted
	Faculty Resources	Ratio of Full-time Science and Clinical Faculty to Full-time Students
National Universities and National Liberal Arts	**Student Selectivity**	SAT or ACT scores Proportion of Freshmen who Graduated in the Top 10 Percent of their High School Classes Acceptance Rate

		Yield
	Faculty Resources	Percentage of Classes with Fewer than 20 Students Percentage of Classes with more than 50 Students Faculty Salary Proportion of Professors with the Highest Degree in their Field Student-Faculty Ratio Proportion of Full-time Faculty
	Alumni Giving Rate	Alumni Giving Rate
Primary Care	**Student Selectivity**	Average Medical College Admission Test Scores Undergraduate Grade Point Average Proportion of Applicants Accepted into the Program
	Faculty Resources	Ratio of Full-time Science and Clinical Faculty to Full-time Students
	Primary-care Rate	Percentage of MDs from a School Entering Primary Care Residencies

Appendix F

Schools with the Biggest Differences in Rank Between 1995 and 2000

Business Schools[a]

School	1995	1996	1997	1998	2000	Negative Change
Brigham Young University	31	45	46	46	49	-18
University of Florida	33	44	38	46	Not Ranked	At least -18
University of Tennessee at Knoxville	37	47	49	43	Not Ranked	At least -14
University of Georgia	40	36	40	39	Not Ranked	At least -10
Pennsylvania State University	37	37	30	31	49	At least -14
School	**1995**	**1996**	**1997**	**1998**	**2000**	**Positive Change**
Arizona State University-Main Campus	44	34	43	31	31	+13
Wake Forest University	46	42	37	39	36	+13
University of Arizona	50	45	35	39	34	+16

[a]In addition to the schools listed above, four schools that were not ranked in the top 50 in 1995 had moved into the top 50 by the 2000 edition and must have jumped at least ten places to do so. These schools were: University of California-Irvine, Rice University, Southern Methodist University and University of California-Davis.

Engineering Schools[a]

School	1995	1996	1997	1998	2000	Negative Change
Case Western Reserve University	30	36	36	41	40	-10
School	**1995**	**1996**	**1997**	**1998**	**2000**	**Positive Change**
Virginia Tech	34	35	32	25	24	+10
University of California-San Diego	43	29	23	21	20	+23

[a]In addition to the schools listed above, the University of New Mexico, which was not ranked in 1995, moved into 40th place in the 2000 edition, and must have jumped at least ten places to do so.

Law Schools[a]

School	1995	1996	1997	1998	2000	Negative Change
Rutgers University School of Law at Newark	41	42	Not Ranked	Not Ranked	Not Ranked	At least -10
University of Oregon	40	Not Ranked	Not Ranked	Not Ranked	Not Ranked	At least -10
School	**1995**	**1996**	**1997**	**1998**	**2000**	**Positive Change**
Brigham Young University	48	32	33	25	29	+19
University of California	45	45	Not Ranked	41	29	+16
Tulane University	50	49	48	45	40	+10
University of North Carolina-Chapel Hill	32	34	35	25	21	+11
University of Notre Dame	39	25	21	25	21	+18

[a]This table does not include schools that were not ranked in 1995 but were in the top 50 in the 2000 rankings. There were six such schools but none would necessarily have had to jump ten ranks in order to get to their 2000 rank (i.e. none of the schools were ranked above 46 in the 2000 edition).

National Liberal Arts Colleges

School	1996	1997	1998	1999	2000	Positive Change
Davidson	21	11	8	11	11	+10

National Universities

School	1996	1997	1998	1999	2000	Negative Change
Rensselaer Polytechnic Institute	39	Not Ranked	48	49	Not Ranked	At least -12
School	**1996**	**1997**	**1998**	**1999**	**2000**	**Positive Change**
University of California-San Diego	43	34	33	32	32	+11
University of Illinois-Urbana-Champaign	45	50	45	42	34	+11