

Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems*

DAVID D. DILL & MAARJA SOO

Research Program on Public Policy for Academic Quality (PPAQ), Department of Public Policy, University of North Carolina at Chapel Hill, Abernethy Hall, Chapel Hill, NC 27599-3435, USA (Phone: +1-(919) 962 - 6848; Fax: +1-(919) 962 - 5824; E-mail: david_dill@unc.edu, maarja@unc.edu)

Abstract. The global expansion of access to higher education has increased demand for information on academic quality and has led to the development of university ranking systems or league tables in many countries of the world. A recent UNESCO/CEPES conference on higher education indicators concluded that cross-national research on these ranking systems could make an important contribution to improving the international market for higher education. The comparison and analysis of national university ranking systems can help address a number of important policy questions. First, is there an emerging international consensus on the measurement of academic quality as reflected in these ranking systems? Second, what impact are the different ranking systems having on university and academic behavior in their respective countries? Finally, are there important public interests that are thus far not reflected in these rankings? If so, is there a needed and appropriate role for public policy in the development and distribution of university ranking systems and what might that role be? This paper explores these questions through a comparative analysis of university rankings in Australia, Canada, the UK, and the US.

Keywords: academic quality, higher education policy, league tables, organizational report cards, university rankings

Introduction

The world-wide expansion of access to higher education has also created an increasing national and global demand for consumer information on academic quality. Because a college education is a rare purchase and an increasingly important as well as expensive decision in one's life, students and their families are seeking information that will help them make informed choices in the selection of a university and/or an academic program. Demand for consumer information on academic quality has led to the development of university rankings in many countries of the world. A UNESCO/CEPES invitational roundtable on rankings and league table methodologies in higher education, for

example, reviewed the development of university rankings in Germany, Japan, Poland, Russia, the UK and the US.¹

The rankings are often heavily criticized: because of their statistical inaccuracy, because of the measures chosen to represent academic quality, or because of their expected negative impact on the overall performance of universities (Bowden 2000). But recent research suggests that well designed organizational report cards can sometimes serve as effective instruments for public accountability (Gormley and Weimer 1999). There also appears to be a growing belief among policymakers, that while various forms of academic quality assurance may be needed to assure academic standards, the provision of relevant information about universities to student consumers is an especially important component of this effort. For example, the government White Paper on higher education in the UK (DfES 2003) argued that market competition could be an important driver of academic quality, if appropriate university information can be provided to help inform student choice.

A comparison and analysis of the existing commercial university rankings or league tables² can help address a number of important questions regarding the rapidly growing international market for higher education. First, is there an emerging international consensus on the measurement of academic quality in these ranking systems, or do important distinctions remain between different countries? Second, applying the criteria used to evaluate other organizational report cards, what are the strengths and weaknesses of commercial university league tables? Finally, should the design of these league tables be left completely to the private sector, or are there important public interests that are thus far not reflected in these rankings? In sum, is there a needed and appropriate role for public policy in the development and distribution of university league tables and if so what might that role be?

This paper will explore these questions through a comparative analysis of the information currently used in commercial university ranking systems in Australia, Canada, the UK, and the US and through a review of related research.

Sample

University league tables may be considered as a type of “organizational report card” that provides explicit organizational rankings (Gormley and Weimer 1999). Such tables have been produced by commercial entities such as newspapers and magazines, professional societies,

Table 1. University league tables examined

Country	Name of the league table examined	Year sampled (regularity)	Web page*
Australia	<i>The Good Universities Guide: 2003 Edition</i>	2002 (annual since 1992)	http://www.thegoodguides.com.au
Canada	<i>The Maclean's Guide to Canadian Universities</i>	2002 (annual since 1991)	Not available
UK	<i>The Times Good University Guide</i>	2002 (annual since 1992)	http://www.thes.co.uk/statistics/main.asp
	<i>The Guardian University Guide</i>	2002 (annual since 1999)	http://education.guardian.co.uk/universityguide/0,10085,488282,00.html
US	<i>US News & World Report America's Best Colleges</i>	2002 (1983, annual since 1987)	http://www.usnews.com/usnews/edu/college/cohome.htm

*Gormley and Weimer (1999) suggest that the "comprehensibility" of an organizational report card, that is its accessibility to consumers, is an important design criterion. They also suggest that web access to organizational report cards may improve comprehensibility. We therefore also evaluated the web sites of the league tables reviewed.

non-governmental organizations (NGOs), as well as by governmental agencies. In this paper we analyze commercially-produced university league tables that have been developed in Australia, Canada, the UK, and the US (Table 1).³ Our focus on the commercial sector reflects the analysis of organizational report cards by Gormley and Weimer (1999), who argue that such rankings can often best be produced by the private sector. Given the increasing emphasis in public policy making on information provision as a means of assuring academic quality in higher education, we wish to test Gormley and Weimers' assertion through a more careful analysis of commercial league tables in the higher education sector.

The paper therefore compares and assesses the following established commercial rankings of first-level higher education: *The Good Universities Guide* (Australia); *The Maclean's Guide to Canadian Universities*; *The Times Good University Guide* (UK); *The Guardian University Guide* (UK); and *US News & World Report, America's Best Colleges*. The paper also reviews relevant literature on the league tables from each country with particular attention to the impacts of the ranking systems on university behavior as well as related research on university choice making decisions among students.

A global definition of academic quality?

The report on the recent UNESCO/CEPES invitational roundtable on rankings (Merisotis 2002) noted that there had been little cross-national analysis of the strengths and weaknesses of university league tables and specifically called for research on whether there are core indicators of academic quality that are consistent across several national rankings. Such insight could also be important given the growing global market of higher education. Over 1.47 million foreign students studied in tertiary education in OECD countries in 1999, a doubling of the number since 1980 (Larsen et al. 2002). These students paid over US \$30 billion in university fees and living expenses to participate in the university programs of their host countries. As this global market for higher education emerges, are commercial university league tables converging on a common definition of academic quality that may ultimately influence the behavior of student consumers and universities around the world?

The five rankings we examine differ in their format, content, and methodology. *Maclean's*, the *U.S. News and World Report* (hereinafter *USNWR*), and *The Times* compose aggregated institutional rankings,

whereas the Australian *Good Universities Guide* (hereinafter *GUG*) and *The Guardian* rank institutions only with respect to particular measures. The *Guardian* ranks academic programs, the *USNWR* and *Macleans* rank institutions, and the *GUG* and *The Times* do both. The *USNWR* and *Macleans* first categorize institutions according to their research/teaching profile, whereas other rankings evaluate all institutions on the same basis. Even the arithmetic is different – *Macleans* ranks institutions with respect to each measure and then aggregates the ranks, whereas the *USNWR* and *The Times* aggregate raw scores. Most importantly, the number and the nature of the measures that the rankings include vary significantly.

In spite of all the differences, however, the rankings suggest a common approach to measuring quality in higher education is emerging internationally. Table 2 compares the measures of the five rankings and divides the measures into input, process, and output measures (see Gormley and Weimer 1999; Pascarella 2001). We can observe that input measures have a prominent role in all five rankings and that the input measures used in the different rankings are quite homogeneous. Process and output measures, on the other hand, are much more diverse and tend to be less influential.

The rankings suggest that one of the leading determinants of a good university is the quality of its incoming students. The academic quality of the student body constitutes 17% of the weight in the *Macleans* ranking, 11% in the *USNWR* formula, and is represented in each of the league tables save the *Guardian*. Quality of incoming students is measured by secondary school grades as well as by university entrance tests. There are several reasons suggested for why the quality of incoming students make a university good or bad. First, the quality of a university may be evaluated by the quality of its output (i.e. its graduates) and measures of the quality of graduates tend to be highly correlated with their ability at entrance. Second, *Macleans* argues that “students are enriched by the input of their peers” and therefore, good entering students are weighted even more in their formula of university quality than, for example, the faculty (20% and 17% respectively). Third, although tautological, it is argued that if a university is able to attract the best students (or international and out-of-province students), then it must be a good university.

The quality of the faculty and research is another prominent shared measure, which is assessed primarily by staff qualifications and the ability to attract research grants. The *USNWR* adds the average faculty salary as an indicator of the “school’s commitment to instruction.” The

Table 2. Performance indicators in five league tables

	<i>GUG</i> (Australia) (by institution and subject)	<i>The Guardian</i> (UK) (by subject)	<i>Maclean's</i> (Canada) (by institution)	<i>The Times</i> (UK) (by institution)	USNWR (US) (by institution)
INPUT	(No overall ranking)	15%*	60%	50%	37%
Faculty	Student/staff ratio % PhDs Research Grants	Student/staff ratio (6%)	% PhDs (3%) National awards (3%) Grants (11%)	Student/staff ratio (9%) Research assessment (14%)	Student/staff ratio (1%) Faculty salary (7%) % PhDs (3%) Full-time faculty (1%) SAT/ACT tests (6%) % top 10% in high school (5%) Acceptance rate (2%) Enrollment rate (2%)
Students	Students ranked in top decile nationally Minimum entrance scores Various measures of student diversity (e.g. # international students, % external, part-time, indigenous, non-English speaking, male/female students)		Avg. high school grades (11%) % top (25%) in high school 3% Out-of-province (1.5%) % International students (1.5%) National academic awards (3%)	Avg. A and AS levels (9%)	

Financial resources and facilities	Non-government earnings	Per student spending (9%)	Per student spending (3%) Student services (4%) Student Scholarships (4%) Library (12%) (# of volumes, volumes per student, % of total budget)	Library and computing spending (9%) Facilities spending (9%)	Per student spending (10%)
PROCESS					
Teaching	Graduate rating of teaching quality and acquisition of generic skills in 30 fields of study	Teaching assessment 65%	Class-size (14%) First year classes taught by tenured or tenure track professors (3%)	Teaching assessment (23%)	Class size (8%)
OUTPUT					
Satisfaction	Graduate satisfaction with courses of study	15%	7%	27%	30%
Graduation			Alumni giving rate (5%) Graduation rate (2%)	Graduation rate (9%)	Alumni giving rate (5%) Graduation rate (16%) Freshmen retention (4%)

Table 2. Performance indicators in five league tables – *Contd.*

	<i>GUG</i> (Australia) (by institution and subject)	<i>The Guardian</i> (UK) (by subject)	<i>Maclean's</i> (Canada) (by institution)	<i>The Times</i> (UK) (by institution)	USNWR (US) (by institution)
“Value-added”		First and upper-second degrees (students who entered with lower grades score more highly) (9%)			Adjusted graduation rate (5%) (controlled for spending and student aptitude)
Learning progress				First and upper second degrees (9%) Job prospects (9%)	
Employment	Job prospects further study Graduate starting salary	Job prospects (6%)			
REPUTATION					
	Student demand, research grant success, and international ranking	Demand among high achieving students	Survey of school guidance counselors, university officials, and organization heads	0%	25%
			Survey of university presidents, provosts, and deans of admission		

* Percentages represent weights assigned to performance indicators in rankings calculation. Percentages may not sum to 100% because of rounding.

student/staff ratio seems to be an important indicator to all but the *Maclean's*.

In contrast to these input measures, assessments of the teaching and learning process seem to get much less attention. In the U.K. the Quality Assurance Agency (QAA) has conducted teaching quality assessments and the results have composed an important part of *The Times* and *The Guardian* rankings.⁴ The only process measure that the *USNWR* employs is class-size. *Maclean's* uses class-size as well as a measure of the exposure of first year students to senior faculty. Two rankings have recognized that the quality of output is highly related to the quality of input and tried to measure the "value-added" of a university. The *USNWR* calculates an adjusted graduation rate that controls for the major input measures of student ability and university expenditures and *The Guardian* measures the proportion of low grade entering students who achieve first or upper second degrees.

While there seems to be an emerging cross-national consensus on input measures indicative of the quality of a university, there appears to be much less consensus on relevant measures of output. The primary output measure utilized by *USNWR* and *Maclean's* is the graduation rate, although its importance varies significantly – 16% of the total score in the *USNWR* ranking compared to 2% in the *Maclean's* ranking. The *GUG* uses graduate employment opportunities as the output measure. *The Times* combines the graduation rate, employment, and learning outcomes. Graduate satisfaction with the academic program is an output measure that is directly measured only in Australia; the *USNWR* and *Maclean's* use alumni giving rate as a proxy for graduate satisfaction.

The reputation of a university is perhaps the most controversial measure. However, with the exception of *The Times* all the rankings seem to believe in the importance of university reputation and two of them are willing to invest a considerable amount of resources attempting to measure that indicator. *Maclean's* conducts a survey among high-school guidance counselors, university officials, the heads of a variety of organizations, and CEOs. *USNWR* asks university presidents, provosts and deans of admission to evaluate other schools' programs and claims to measure the "intangible" aspects of learning and teaching in that way. In the *GUG* reputation is described as "prestige" and is measured by student demand, success in attracting research grants, and success in international ratings. *The Guardian* also measures reputation by student demand, a high-score being based on attracting students with good A-levels. Although the measures vary, the concept of reputation has a prominent place in university rankings.

In sum, the comparison of the five national league tables shows an emerging international consensus on the definition and measurement of the academic quality of first-level degree programs. The producers of these commercial league tables suggest academic quality can be assessed primarily by input measures and by academic reputation. Input measures include the quality of enrolling students, the quality of the faculty, and the financial resources available to a university. Measures of teaching and student learning processes are generally unavailable, with the exception of the UK where academic programs have been independently assessed. Output measures are also limited, with an emphasis on graduation rates, employment prospects of graduates, and alumni satisfaction.

What are the strengths and weaknesses of these commercial efforts to provide consumer information on academic quality and what are the possible influences of these rankings on the emerging global market for higher education? We will pursue these questions by evaluating the five league tables with a set of criteria developed for assessing organizational report cards (Gormley and Weimer 1999).

Evaluating the league tables

Organizational rankings can sometimes serve as a useful instrument for public accountability, supplying information to both consumers and policy makers on measurable differences in service quality, while also providing an incentive to organizations for quality improvement (Gormley and Weimer 1999). Whether rankings or league tables make such a contribution to the public interest depends upon how they are devised. Critical criteria identified for evaluating the design of effective organizational report cards include the *validity* of the measures, the *comprehensiveness* of the measures, the *relevance* as well as *comprehensibility* of the information provided to student consumers, and the *functionality* of the rankings in motivating improvements in teaching and student learning within organizations (Gormley and Weimer 1999).⁵

Validity

A critical criterion for the design of effective university report cards is validity. Gormley and Weimer (1999) argue that the validity of organizational report cards should be evaluated on two dimensions.

First, does the report card focus on measures that closely approximate or are clearly linked to valued societal outcomes, which in the case of first-level university degree programs are the knowledge, skills, and abilities achieved by graduates?⁶ Second, since report cards are designed to compare the performance of organizations, it is also important for rankings to control for differences between organizations in client characteristics and resources in order to detect the actual marginal contribution made by the organization itself. In other words, well-designed report cards will attempt to measure the “value-added” by an organization. In the case of university report cards this would require controlling output information on graduates by entering student ability in order to identify the contribution directly attributable to the quality of education provided. Whether report cards also controlled for resources would depend on the intended audience. Students would like to know the educational “value-added” relative to their own private costs, but may be unconcerned with the overall costs to society. Policymakers or regulatory agencies on the other hand would be very interested in evaluating the educational “value-added” relative to overall resources.

The most effective means of maximizing the validity of a university report card is to include measures known to be associated with student outcomes valued by society. We assume that socially valued student outcomes are those that contribute to human capital (Becker 1964). During their university education students’ develop knowledge, skills, and abilities that over their lifetimes provide private benefits to themselves as well as social benefits or social capital to the larger society. This human capital perspective provides the logic for public subsidies for higher education and is also explicitly reflected in current national policies on academic quality, which seek to improve the academic standards of higher education institutions (Brennan and Shah 2000). Consistent with human capital theory these policies increasingly focus on information about student learning outcomes, that is, the educational “value-added” of an academic program or degree (Dill 2000).⁷

Research on the impact of universities on students suggests certain conditions, within the control of academics, that increase student learning (Pascarella and Terenzini 1991). Factors such as the nature and cohesiveness of students’ curricular experiences, their course taking patterns, the extent to which faculty members involve students actively in the teaching-learning process, non-classroom interaction with faculty members, and the amount of peer group interaction have all been discovered to be associated with student learning. A longitudinal survey by Astin (1985; 1996) similarly reveals that the learning environment and

student involvement in the learning process have the main impact on students' cognitive and affective outcomes. But it is precisely these types of process measures that, along with measures of university outputs, are missing in many of the reviewed league tables. As Pascarella (2001) recently concluded:

A more serious problem with the national magazine rankings is that from a research point of view, they are largely invalid. That is, they are based on institutional resources and reputation dimensions, which have only minimal relevance to what we know about the impact of college on students (p. 20).

Indeed, the five reviewed league tables fail to provide a theoretical or empirical justification for the measures selected and the weights utilized to calculate their rankings. From the measures utilized in the tables we would infer that prominent research institutions give the best education, although it is more accurate to conclude that the listed performance indicators do a much better job in assessing the research quality of a university than its teaching quality (Yorke 1998).⁸ The rankings are heavily biased toward measures known to be associated with research performance: financial resources, numbers of faculty and research activity, student selectivity, as well as university reputation. Even the average faculty salary, which according to *USNWR* measures a school's commitment to instruction, more likely reflects faculty orientation to research and has been found to be negatively correlated with student learning (Astin 1996).

Reputation is an important component in the rankings. *USNWR* claims that their peer assessment of reputation is aimed at measuring "intangibles" such as faculty dedication to "teaching." The *USNWR* reputation score, however, correlates much more closely with high per-faculty federal research and development expenditures than with good graduation-rate performance (Graham and Thompson 2001). Astin (1985) points out that there is in fact no need for expensive surveys of deans and presidents to identify academic reputation, because the *USNWR* reputation measures can be quite well predicted by three objective and readily available indicators: undergraduate selectivity, per student expenditure, and number of doctorate-granting departments (see also the related research results of Grunig 1997 and Paulsen 1990).⁹ Another problem with the *USNWR* reputation survey is that while it may be appropriate for ranking the best known schools, even a sample of prominent people is not able to assess accurately the quality of all programs in all schools. Therefore their opinion is likely to be influenced

more by the existing reputation of the university (i.e. “halo effect”) than by actual knowledge of program quality (Clarke 2002).

But perhaps universities with strong research orientation have the best learning environment and give the best education? Empirical research, however, suggests that the correlation between research productivity and undergraduate instruction is very small and teaching and research appear to be more or less independent activities (Terenzini and Pascarella 1994). Astin’s (1996) studies specifically explore the nature of the relationship between research and teaching in the US. A department that has a strong research orientation (i.e. a department that publishes many books and articles, spends a substantial amount of time on research, and attaches high personal priority to engaging in research) has a negative correlation with factors having to do with teaching: hours spent teaching and advising, commitment to student development, use of active learning techniques in the classroom, and the percentage of faculty engaged in teaching general education courses. In addition, research orientation has a negative effect on student satisfaction with faculty as well as on student’s leadership, public-speaking, and interpersonal skills (Astin 1996).¹⁰

The rankings also place heavy weight on input measures, although empirical studies show that most input indicators have an irrelevant or very small effect on students’ learning. After reviewing over twenty years of empirical research on the impact of college on students Terenzini and Pascarella (1994) concluded that the supposed influence of inputs on student learning was one of the great myths of higher education. That is, after taking into account the characteristics, abilities, and backgrounds they bring with them to college, how much students grow or change has only an inconsistent or trivial relationship with such input measures as educational expenditures per student, student/faculty ratios, faculty salaries, percentage of faculty with the highest degree in their field, research productivity, size of the library, admission selectivity, or prestige rankings.

As suggested in the earlier quote from *Maclean’s*, there is also a strong belief in “peer effects” in higher education. That is that the overall quality of students entering a university has an independent influence on graduates’ success. While there is evidence to support the influence of peer interaction on student learning (Pascarella and Terenzini 1991), the positive academic benefits of peers are obviously dependent to some extent on the nature of the university education. For example, the extent to which the university’s processes for teaching systematically encourage student interaction on academic tasks. However, serious questions need

to be raised about the assumed positive relationship between peer effects, as measured by average entering student test scores, and human capital formation. Empirical research in support of this relationship is based largely on econometric studies of the relationship between average entering student test scores and graduate lifetime earnings as well as a small number of studies of the effects of peer quality (again as measured by entering test scores of freshman roommates) on grade point averages in US colleges.¹¹ In contrast, the extensive research on student learning indicates an inconsistent and trivial relationship between admissions selectivity based upon average entering student test scores and measures of the knowledge, skills, and abilities learned by students during their education (Pascarella and Terenzini 1991).

The most recent review of the peer effects research also casts significant doubt on the supposed relationship between peer effects, as measured by average test scores of entering students, and students' earnings capabilities.¹² First, the research notes that the impact of institutional selectivity on earnings is nonlinear. Only the most selective institutions may have an impact on earnings. Second, the relationship depends on the students' major field of study, which is often not controlled in relevant studies. That is, less selective, public institutions in the US often offer academic majors with less potential earnings capacity than selective schools. Finally, and most importantly, there is an indication that if researchers control for the types of students who apply to more selective institutions – utilizing measures of individual ambition – the earnings advantage of more selective schools disappears.

The belief in “peer effects” has contributed to the prominence in league tables of measures of the quality of entering students such as average student test scores. But even if these effects exist, they are likely limited to a small number of institutions and vary by academic program within each university. Ranking all universities using measures of student selectivity based upon median entering test scores therefore provides information of little value to the majority of university applicants. Furthermore, if university league tables are intended to offer potential students information on the earnings and/or employment benefits of attending particular universities, then rankings of whole institutions are seriously misleading. The information needed by student consumers is on the performance of specific subjects or programs. Finally, as we will discuss below, the focus on student selectivity as measured by entering student test scores tends to distort institutional behavior by providing incentives for all universities to focus scarce resources and administrative

effort on improving student selectivity rather than investing in academic quality improvements for the students enrolled (Ehrenberg 2003).

If the validity of input measures is questionable, then potentially output measures offer better indicators. Output measures utilized in the rankings include graduation rates, graduate satisfaction, and graduate employment. However, while the number of students who graduate from university is certainly a socially valued outcome, the fact that graduation rate can be independently controlled by each university poses a problem. That is, graduation rates can be increased both by more effective teaching and student learning and by lowering academic standards. The issue of university grade inflation and inflation in honors degree awards recently has been raised in both the US and UK (Rosovsky and Hartley 2002; Yorke et al. 2002). Alternative output measures as well as additional quality assurance mechanisms may therefore be needed to assure that league tables and competitive markets do not create incentives for dysfunctional responses on the part of universities.

Graduate employment measures are attractive output measures, but these indicators are still vulnerable to criticism. Employment information utilized in the UK league tables reports the proportion of students that have found a job six months after graduation without controlling for the individual's social class background, class of degree, the degree subject studied, or local labor market conditions all of which have been discovered to influence scores (Smith et al. 2000). Neither does it identify whether students are employed on graduate level jobs or are under-employed. Analysis of the UK data (Smith et al. 2000) also suggests that there is no statistically significant difference between most UK universities in the pattern of graduate employment; perhaps only the top 10 and bottom 10 universities have a meaningful difference in their results.¹³

Astin notes that "it is difficult to argue that any other outcome category - cognitive or affective - should be given greater priority than student satisfaction" (Astin 1991, p. 62). The *GUG* utilizes the results of a government mandated Course Experience Questionnaire (CEQ), which measures how satisfied graduates are with their program.¹⁴ *Maclean's* and *USNWR* use alumni giving rate as an indicator of graduate satisfaction, but this measure may be more a function of the vigor of the development office and the tradition of fund raising at that institution than a measure of student satisfaction (Ehrenberg 2002b). For example, the measurement of alumni giving rate has encouraged universities to adopt more aggressive tactics to get small contributions

from more alumni. As an alternative measure, many US institutions regularly ask graduating seniors to complete satisfaction surveys. If universities were required to publicly provide such information commercial university league tables would thereby have access to a much more valid indicator of student satisfaction than alumni giving rate (Graham and Thompson 1994).

Another limitation of the existing outcome measures (e.g. graduate employment opportunities, graduation rate, graduate satisfaction) is that they may reflect the universities' recruitment policies instead of the actual quality of education. Rankings have started to pay attention to this limitation. In 1996, *USNWR* introduced a "value-added" measure that provides a graduation rate for each university controlled for university spending and entering student test scores. The construction of this value-added measure reflects Gormley and Weimer's (1999) stated concern about the need to control for critical inputs in order to effectively evaluate the performance of an organization. *The Guardian* also includes a value-added measure, which controls students' degree performance by their entry standards only.¹⁵ In the case of *USNWR*, as Table 2 indicates, the value-added measure contributes only 5% to the overall ranking. Furthermore, faculty salaries and per student spending independently contribute 17% and relevant characteristics of the entering student body including test scores and class rank contribute another 11%. Thus the amount of "correction" provided by the *USNWR* value-added measure, as well as the influence of the measure itself, are effectively compromised in the overall ranking.

Finally, there have been numerous analyses of the relevant league tables that have raised questions about their construct validity (Clarke 2002; Eccles 2002; Morrison et al. 1995; National Opinion Research Center 1997; Page 1999; Yorke 1997). These include the extent to which the various performance indicators are in fact measuring relevant factors or dimensions of academic quality, whether there is a statistically defensible rationale for the weightings employed in the various league tables, and the legitimacy of constructing rankings when there are not statistically significant differences in the institutional data. That is, some of the universities ranked lower in the league tables do not differ in any meaningful way from those institutions ranked higher.

Based upon this overall review of the validity of university league tables we would argue that more valid university rankings would have the following characteristics. First, they would focus on process measures that research has demonstrated to be clearly linked to student learning as well as relevant student output measures. Input measures of

faculty, students, and resources would be given minimal weight and would be used primarily as controls on relevant output measures. This can be done through value-added measures such as that developed by *USNWR* or less rigorously by grouping institutions into different categories according to inputs (e.g. student ability or resources per student) or mission as is now done in both the *USNWR* and *Macleans*'s rankings.¹⁶ Reputational measures, particularly those based upon surveys, would be given little weight. The table would provide process and output information by academic subjects and programs, although relevant process measures on the institution itself could be valuable, especially in countries where first degree education is less specialized. Finally, rather than pretending that differences between ranked institutions provide statistically meaningful information, universities would be ranked alphabetically within hierarchically ranked categories similar to the format utilized in various consumer guides.

Comprehensiveness

A second criterion is comprehensiveness: does the report card employ a range of indicators that capture the critical dimensions of academic quality? Given the complexity of both the outputs of first degree level academic programs and the processes of teaching and learning it is important that the measures utilized in league tables be comprehensive so as not to produce rankings that are incomplete or misleading.

As previously discussed Table 2 summarizes the performance indicators used in the five reviewed league tables and classifies them into input, process, and output measures. The most comprehensive measurement, as illustrated by the number and variety of indicators, is on faculty, student and financial inputs. Reputation, which we have argued is a less valid and relevant indicator for a university report card, is measured with surveys by both *USNWR* and *Macleans*'s, with predictable concerns about reliability and validity depending on the design of the survey instrument, the knowledge of those surveyed, and the response rate. In contrast the *GUG* uses multiple objective indicators to assess reputation and *The Guardian* uses an objective measure of admissions selectivity as a proxy for reputation.

Processes, which we argue should receive a primary emphasis in valid university report cards, are not comprehensively assessed in most of the league tables. Indeed in North America only class size and in Canada first year classes taught by tenured faculty are utilized in the rankings.

Valid process measures of academic quality do exist in the US. The National Survey of Student Engagement (NSSE) can provide information on how effectively colleges are contributing to learning in five areas: level of academic challenge; active and collaborative learning; student/faculty interaction; enriching educational experiences; and supportive campus environment (Kuh 2003). The indicators utilized in the NSSE are derived from extensive research on factors related to effective student learning in colleges and universities. *USNWR* provides this survey data on their website, but only for colleges and universities that have given their agreement to do so. With the exception of private Rice University and the public universities of Michigan and University of North Carolina at Chapel Hill, none of the other 50 “best” national universities permitted their data to appear in the 2002 *USNWR*.¹⁷ As a consequence these informative process indicators are not included in *USNWR*’s published college and university rankings.

It is difficult to argue that class size captures the critical dimensions of curricula and teaching that universities may use to improve the quality of teaching and student learning. Indeed, by not incorporating the NSSE data on learning processes into its rankings, *USNWR* may be encouraging universities to take simplistic process actions to enhance their ratings rather than engage in the more challenging task of improving teaching and student learning.

In strong contrast to the US and Canada the UK league tables place a great deal of emphasis on Teaching Quality Assessments (TQA). While TQA is presented as a single process indicator in the UK league tables, these scores are in fact based upon direct observations and comprehensive assessments of the curricula and teaching behaviors in university subject fields and also include a number of objective measures of the academic environment of each university.¹⁸

Finally, in Australia and the UK graduate job prospects offer an additional indicator of output not currently utilized in league tables in the US or Canada. While as noted there are limitations with this measure as well, the potential combination of graduation rates, graduate satisfaction, and job prospects measures suggests the possible strengths of utilizing a more comprehensive set of output indicators.

As we have noted above, the validity and reliability of a number of the performance indicators of university quality currently used in the sampled university league tables are debatable. Survey data, both for measures of reputation and for graduate satisfaction, have a number of known limitations and in the latter case have been manipulated by some US institutions to inflate their rankings (Ehrenberg 2002b). Output

measures, such as retention or graduation rates, which are directly controlled by the institution, could conceivably be increased by lowering academic standards. Proxy indicators of graduate satisfaction, such as alumni giving, may tap a broad range of student experience and poorly represent student satisfaction with academic programs. A process indicator such as class size may be too simplistic a measure for large, complex universities. For these reasons effective league tables would need to utilize a comprehensive set of relevant input, process, and output indicators and employ measures that utilize different sources and types of data.

Relevance

A third criterion is relevance. Does the report card present information relevant to the needs of student consumers – for example does it provide information appropriate to the specific choices students must make? One indirect measure of the consumer relevance of university report cards is the nature of the readership or purchasers of these league tables. Research in the UK and US suggests that commercial league tables are most often designed for and used by a narrow segment of the potential student market – students of high achievement and social class (Carrico et al. 1997; McDonough et al. 1998). Many of these students appear interested in the “prestige” rating of a university as reflected in the future opportunities and incomes of an institution’s graduates. But research in Australia, the UK, and US on the preferences of student applicants also suggests that league tables that provide university rankings based upon a single weighting scheme do not meet the needs of the majority of students, who desire a much more varied list of factors for deciding where to apply (Carrico et al. 1997; Connor et al. 1999; James et al. 1999; McDonough et al. 1998; Moogan et al. 1999). As Ehrenberg (2002b) concluded with regard the *USNWR* rankings:

Indeed, once one realizes that different students may value the characteristics of universities differently, the notion that one can come up with a single number that summarizes the overall ranking of an academic institution seems quite silly. (p. 53)

For example, a recent survey on student choice in the UK (Connor et al. 1999) indicates that the most important factors influencing the choices of applicants to full-time university education are the course or subject, academic quality (particularly teaching reputation), entry

requirements, employment prospects for graduates, location, available academic and support facilities, social life, and costs of study. Despite the differing structure of American higher education, the extensive US research on college choice suggests that similar factors are important for US students and parents in choosing among colleges. The most significant factors include the academic program (major area of study), tuition costs, financial aid availability, general academic reputation/general quality of institution, location (distance from home), college size, and social atmosphere (Hossler et al. 1989; Manski and Wise 1983; Paulsen 1990; Zemsky and Oedel 1983).

Information on the academic subject has consistently proven the most influential on student choice in Australia and the UK (James et al. 1999; Moogan et al. 1999) and raises fundamental questions about the utility of league tables that provide rankings and information only for the overall university. First, highly ranked universities may not have the specific subjects sought by a student. Second, entry qualifications may vary across subject fields even within the same university.¹⁹ Finally, and most importantly, the quality of the student learning experience, graduation rates, student satisfaction, employment prospects, and even lifetime earnings are apt to vary significantly by subject field within the same university. Therefore, rankings based upon average data for the university as a whole not only misrepresent the experience for particular subject fields, but fail to provide the type of academic quality information most desired by student consumers.²⁰

In sum, a league table will have more relevance for student consumers if it utilizes performance indicators and provides information that focuses on the quality of teaching and student learning, the experiences and structure of academic subjects or courses of study, and opportunities for graduates.

Comprehensibility

A fourth criterion is comprehensibility. Does for example the amount and form of information provided by the report card and the media by which it is transmitted meet the needs of student consumers? While market research suggests that the more complex the information provided the poorer the resulting consumer choice, studies of school report cards have noted that parents prefer more detailed information as well as formats that permit them greater control over the type of information they receive (Gormley and Weimer 1999). The extent of information

desired varies by both income and education; arguably university report cards might therefore provide longer and more technical reports.

Gormley and Weimer (1999) provide some specific suggestions on means to make report cards more accessible for consumers. For example, how the information is presented and explained is likely to affect the comprehensibility of the rankings. Ratings (below average/average/above average or good/satisfactory) are likely to be easier for consumers to interpret than are raw scores. Similarly using ordinal level rather than interval level data is likely to be more consumer friendly. In addition to the ranking information, a readable summary and a clear explanation of the ranking methodology are likely to contribute to the comprehensibility of report cards.

Consumers also desire information that they believe is relevant and corresponds to their specific needs and expectations. The recent UNESCO/CEPES Report (Merisotis 2002) emphasized that league tables will need to provide consumers more control and ownership over what is actually being ranked. As access to higher education continues to expand and the demographics of student applicants become increasingly diverse (Pascarella and Terenzini 1998) the ability of student consumers to obtain information on universities relevant to their particular preferences becomes even more crucial. The Internet provides a possible opportunity to develop a personalized approach to report cards.

Internet-based information on university report cards has potential appeal because of its low marginal cost of access, easy discovery through commonly available search engines, and ability to provide information that corresponds to the specific needs and expectations of the consumer. Market research suggests that consumers appreciate access to precise, detailed information that answers their questions rather than extensive amounts of information that does not meet their needs (Gormley and Weimer 1999). The UK survey of student choice (Connor et al. 1999) indicated that Internet-based information was used less frequently than print-based sources and this difference was also true for mature applicants. It is also likely however that interest in and access to the Internet will become more widespread in the years to come.²¹

A more comprehensible league table would therefore, provide a wide variety of relevant information for student consumers and offer Internet access to the information that would permit them to define and conduct their own search according to their individual priorities. As Bowden (2000, p. 58) notes, students would then be able to “find an answer not to the question ‘which is the best university?’, but to the much more appropriate question ‘which is the best university course for me?’”

Functionality

The final criterion is functionality. Is the report card designed in a way that encourages the ranked universities to engage in the improvement of teaching and student learning, or does it create incentives for dysfunctional university behavior such as data misrepresentation or student recruitment designed to inflate ranking scores (i.e. “cream skimming”)?

Although university league tables are aimed by intention and design at consumers, the logic of this market-based approach to quality assurance is that the choices of student consumers will eventually affect the academic quality of the universities themselves (Tight 2000). Given the public nature of these report cards university leaders may also anticipate the effects and act to respond in ways that will advance the interests of the institution. How have universities responded to the development of these league tables?²²

The earliest reports in the US suggested some colleges and universities were manipulating data central to the league table rankings, for example attempting to increase their average entering student test scores by dropping out the lowest scores or not reporting the scores of international students (Steklow 1995). This type of manipulation is more possible in the US and Canada, because much of the data used in university league tables is self-reported by the institutions themselves and is not subject to government regulation or audit.

More recent reports suggest that a number of US universities are attempting to “game” the commercial university rankings to better position their institution on the measures employed (Ehrenberg 2002b). For example, because the proportion of living alumni who make contributions to an institution is used by *USNWR* as a proxy for graduate satisfaction, Cornell University administrators lowered their count of alumni reported by eliminating those for whom they did not have good addresses and those who had attended but not graduated from the university. Because *USNWR* uses total expenditures per student (rather than relevant educational expenditures per student) as a proxy for academic quality, Cornell administrators also realized they could enhance the university’s rankings by including expenditures for Cornell University Medical College in the University’s league table submission even though this is a wholly post-graduate college and is located in New York City hundreds of miles from the main campus in Ithaca.²³ Finally, several institutions have recently made the Scholastic Aptitude Test (SAT) an optional requirement for applicants (Ehrenberg 2002a). This was publicly justified as providing opportunities for a more diverse

pool of university applicants, but a number of these institutions were known to be dissatisfied with their *USNWR* ranking. Because of the specific measures used in the *USNWR* league table, making the SAT tests optional will likely raise the institution's rankings. That is, only students with high SAT scores are likely to report them and applicants with lower test scores will now more likely apply. Therefore the colleges should be able to increase the average test scores of entering freshmen and lower the fraction of freshmen applicants admitted, both of which are influential measures in the *USNWR* rankings.

A number of colleges and universities have also adopted early admissions plans for students who will make a commitment to a particular institution. Because almost all early applicants eventually enroll, such programs lower the fraction of total freshmen applicants that need to be admitted and also increase the institutions' "yield" rate, both of which were given substantial weight in *USNWR* league tables (Ehrenberg 2002a, 2002b; Monks and Ehrenberg 1999).²⁴ What is conspicuously missing in all these reports of college and university responses to US league tables are active efforts to improve teaching and learning for students.

In a recent national study of US colleges and universities Rand researchers (Brewer et al. 2002) detected evidence of an increasingly costly "arms race" for prestige among large numbers of colleges and universities.²⁵ Many institutions are making extensive investments designed to increase the selectivity of the admissions process by linking tuition discounts with academic merit and student ability, attempting to lower student acceptance/yield rates, and investing in student consumption benefits such as dormitories, eating facilities, or fiber optic computer networks that will help attract high ability students. The Rand researchers suggest that this attempt to build prestige by "cream skimming" the student market does not seem to lead to an improvement in the quality of educational delivery and may lessen the overall educational benefits of higher education for students and ultimately for society. The Rand researchers suggest that this pursuit of prestige through increasingly costly investments in admissions selectivity is reinforced by commercial college ranking systems in the US that use student "inputs" as a primary measure in national league tables.

Several case study reports suggest an interesting and different response of US universities to league tables of research doctoral programs as compared to their response to league tables of first-level academic programs. Trow (1983, 1999) describes the extensive changes made in the departmental structures of the biological sciences at the University of California, Berkeley, the means of appointing and

promoting faculty members in the university's biological community, and in the nature of facilities for doing biological sciences over a twenty year period. He argues that the impetus for these dramatic changes came in part from the decline in the rankings of several of the biological sciences departments revealed in the National Research Council assessment of research doctoral programs in 1982. Similarly Ehrenberg (Ehrenberg and Hurst 1996; Ehrenberg 2002b) discusses how administrators at Cornell University utilized rankings and data from the National Research Council quality assessment of 1993 to improve the Department of Sociology and Biology programs. In Sociology the analysis revealed that the department's low ranking was due to its small size not to its faculty's productivity, therefore, the university decided to continue the department and increase its number of faculty. In biology the assessment led the university to devote resources to particular areas in which the university had special strengths and which would likely be important in the coming years. The university also prodded the biology programs to build better links between their activities on the Ithaca campus and the biology programs in the Cornell Medical School in New York City. Similar improvements in the management of university research and in the reorganization of research units have been reported as a consequence of the Research Assessment Exercise (RAE) in the UK.²⁶

How do we account for this apparently different response of universities to league tables of research-doctoral programs and league tables of the quality of first level academic programs? Ehrenberg (2002b) notes that the National Research Council rankings, while based upon subjective peer judgments, also include objective data on a number of important measures of research-doctoral programs. These include inputs such as the number of faculty members and doctoral students in each program, and process measures such as student time to degree. The measures also include outputs such as number of doctoral graduates each year and number of faculty publications, as well as significant outcomes, such as the number of times faculty publications were cited and the number of distinguished awards received by the faculty. Ehrenberg and Hurst (1996) were able to use these objective measures to develop a causal model of "prestige" in research doctoral programs that helped to guide the strategic decisions made by administrators at Cornell. We would argue that it is this lack of a demonstrated causal logic between the measures used in many of the league tables for first-level degree programs and student learning outcomes that encourages the gaming behavior and strategic mis-investments reported in the US.²⁷

In sum, league tables with greater functionality will be designed to encourage universities to make improvements in teaching and student learning. Less functional league tables will provide incentives for universities to “game the system” through manipulation of data and investments in costly expenditures designed to “cream skim” student applicants and enhance the university’s overall reputation. We believe that the functionality of university league tables is to some extent influenced by regulations affecting the reporting of university performance data as well as the public availability of relevant information on university processes and outputs. In a following section we suggest some government actions that could improve both the relevance and functionality of university league tables.

A report card on university league tables

Given this review of the criteria important for designing effective organizational report cards, how do we assess the five university league tables? That is, how do the league tables compare on validity, comprehensiveness, comprehensibility, relevance, and functionality – what are their relative strengths and weaknesses?

Table 3 presents our report card on the five league tables based upon the five criteria. As way of introduction, this simple exercise clearly illustrates a number of the obvious issues with league tables generally. First, we have attempted to justify the validity of our five performance indicators or criteria by basing them upon factors discovered to be important in the design of effective organizational report cards in a wide variety of fields (Gormley and Weimer 1999). We have further refined

Table 3. A report card on university league tables*

	Validity	Comprehensiveness	Comprehensibility	Relevance	Functionality
<i>GUG</i>	XX	XXX	XXX	XXX	XX
<i>Guardian</i>	XX	XX	XX	XX	XX
<i>Maclean's</i>	X	X	X	X	O
<i>Times</i>	X*	XX	XX	X	X
<i>USNWR</i>	X	X	XX	X	O

Scores. XXX: very good; XX: good; X: adequate; O: inadequate
SOURCE: Authors' judgment!

*Adapted from Gormley and Weimer (1999, p. 226).

these criteria through a review of relevant research in higher education. It is worth noting that none of the five reviewed league tables provides a similar theoretical or empirical rationale for their choice of measures. Furthermore, our assessment of the five league tables on each of the five criteria is based upon our personal judgment. Similarly, the basis for the institutional rankings on a number of the indicators in these university league tables is also subjective. Finally, we provide a comparative report card on the university league tables in alphabetical order, not a ranking. Therefore, unlike the university league tables themselves, we are not obligated to justify a weighting of the different criteria presented or determine whether there are statistically meaningful differences between the five tables assessed.

In reviewing the five league tables we made the following judgments in accord with our stated criteria. We rate the validity of *Maclean's*, and *USNWR* as barely adequate because of their heavy reliance on subjective rankings of reputation and input measures, as well as their inadequate measures of process and outputs. Both league tables also compute whole university rankings using the same criteria for each institution, although to their credit both rankings group institutions with comparable peers. *USNWR* also has developed a measure of value-added, but gives it too little weight. *The Times* validity is also compromised to some extent by ranking of whole institutions on the same criteria and the emphasis on inputs, particularly Research Assessment ratings. We judge *The Times* as more valid (i.e. adequate-starred) than its North American peers because of the inclusion of Teaching Quality Assessment data and multiple measures of output including graduate employment prospects. We believe that both the *GUG* and *The Guardian* possesses good validity, but for different reasons. Both do not offer a single ranking of institutions. The *GUG* ranks institutions in divisions or bands according to a variety of criteria, with no overall ranking for all institutions. The *GUG* also provides graduate ratings of the educational experience in particular programs, including measures of overall student satisfaction, teaching quality, and generic skills. *The Guardian* provides rankings only by academic program and places the greatest weight in its rankings on the TQA (process) and a value-added measure of honors degrees awarded that controls for entering student ability. Reputation, which is measured objectively, and inputs are given relatively little weight in *The Guardian* rankings.

Both *USNWR* and *Maclean's* have only adequate comprehensiveness of measures. Both use extensive numbers of indicators for inputs, but many of these are self-reported by the institutions and not

independently verified.²⁸ They both rely on class size data as a process measure, as well as alumni giving rate and graduation rate as output indicators, despite the potential for institutional manipulation of these latter measures. Finally, both put an extraordinary weight on reputation measures derived from surveys. In contrast, both *The Times* and *The Guardian* use multiple measures of outputs, including graduate employment prospects, and the majority of their data sources are objective indicators derived primarily from government records. The *GUG* has multiple measures for almost all its dimensions including reputation, which is objectively measured using three indicators. The source of almost all its data is government records.

With regard the comprehensibility of the league tables to consumers, we judge the *GUG* very good and the other rankings, save *Maclean's*, to be good, although they each have different strengths. The *GUG* makes very effective use of ratings. The percentage of graduate employment and starting salaries, for example, are given numerically as well as on an "average-better-worse" scale. The *GUG* also uses a simple five-star system to group universities. *The Guardian* uses ordinal measures from 0 to 6 for student/staff ratio and reputation, and a 0–10 scale for spending per student. *The Times*, *Maclean's*, and *USNWR* provides raw scores only. *The Times* ranking, for example, gives the percentage of unemployment in decimal points. *Maclean's* presents a rank instead of a performance score in its main table, but raw scores are presented separately at the end of the guide. All the league tables studied provide at least some information about their methodology, although they do not offer much explanation about their choice of criteria. *Maclean's* is an exception in this respect and provides a description of each criterion utilized in their ranking methodology.

As noted in Table 1 all of the league tables reviewed, except for *Maclean's*, provide a website with access to their information on university rankings. For this reason we do not judge *Maclean's* to be as comprehensible as the other league tables. *The Guardian* provides free access; for the other league tables there is an extra charge to users of the website. *The Guardian* and *USNWR* websites offer interactive, web-based versions of their ranking systems, which we believe to be a valuable aid to consumers who can thereby rank universities on criteria of importance to them. *The Guardian* website for example allows the user to assign different weights on criteria and by this means to construct a personalized ranking. The *USNWR* website permits the consumer to identify universities with specific characteristics (e.g.

a combination of selectivity, location, extracurricular activities, etc.). *The Times* and the *GUG* websites permit consumers to construct a ranking according to the criteria available in their paper issue and therefore do not provide the consumer with more personalized options. However, as noted above, the comprehensiveness of the criteria and measures included in the *GUG*, and, as will be discussed below, their relevance to student consumers, suggests their website is likely to provide effective personalized searches.

On the criterion of relevance of the information presented to student consumers we give a very good mark to the *GUG*, because of its broad array of indicators, data available from the Graduate Destination Survey and Course Experience Questionnaire, and its focus on subjects and academic programs. *GUG* presents the most complete collection of information relevant to the preferences expressed by student consumers and the variety of information available allows students to individualize their searches in order to meet their particular needs. While *The Guardian* provides much less of the type of information requested by students it does focus on providing information on teaching and on the quality of subjects or programs, both of which as noted above are a primary concern for most student consumers.

Finally, the fact that the majority of the information utilized in the league tables in Australia and the UK is derived from government agencies seems to have led to less dysfunctional data manipulation and “gaming of the system” than has been reported in the US. It also appears that the publication of data derived from the government mandated Graduate Destination Surveys and Course Evaluation Questionnaires in Australia and the Teaching Quality Assessments in the UK may have provided some incentive for institutional improvements in teaching and student learning in these countries.²⁹ While we believe even more valid league tables would provide greater incentives for improvement in university quality, we rank the *GUG* and *The Guardian* as good in functionality. *The Times* is ranked acceptable, because of its continued emphasis on research assessments. In comparison, the reported emphasis on “cream-skimming” high achieving applicants and extensive institutional expenditures designed to increase reputation in the US suggests that the design of the *USNWR* league table and the similarly constructed *Maclean’s* league table are inadequate on functionality. As will be discussed below these types of university league tables may be providing incentives for university activities that do not increase the socially valued outcomes of higher education.

Role for government?

A case has been made that provision of information through organizational report cards is a more efficient and effective means of achieving the public interest in government provided or subsidized services than direct regulation (Gormley and Weimer 1999). However, our review of five national university league tables suggests that these expected benefits are unlikely to occur if report cards are designed primarily to assess the overall reputation of universities. The commercial provision of such “beauty contests” to those who wish to purchase them may be relatively harmless, assuming more valid, relevant, and functional information is readily available to the broader public. However, a number of observers (Brewer et al. 2002; Ehrenberg 2002b) have argued that in the US, despite the large number of guides and handbooks currently available, college and university rankings may be an important contributor to a socially inefficient “academic arms race” in higher education. League tables such as *USNWR*, while used primarily by a select group of students, may shape public opinion about what constitutes a quality education in ways that negatively affect both student consumers and institutional behavior. For example, information on academic programs relevant to student consumer choice, such as that readily available in the *Good University Guide* in Australia, is still not currently available in the US. Further, there is some evidence that the focus of US league tables on reputation and the particular indicators used to measure reputation may be contributing to the continually rising costs of US higher education as well as providing incentives for colleges and universities to invest in actions and strategies that actually detract from the social benefits traditionally provided by higher education. To the extent such circumstances exist they may warrant government intervention to assure that appropriate consumer information is provided on higher education, actions similar to those taken for other consumer products and services that are deemed of significant importance to the public interest.

One reason for our judgment as to the superiority of the Australian and UK league tables over those available in North America is that public policy in the former countries requires that universities provide information relevant to the design of effective report cards. Thus in our judgment league tables in Australia and the UK are more reliable than those in North America because they are based primarily upon objective data on university performance collected systematically by government agencies (Eccles 2002). In comparison, data used for league tables in Canada and the US rely heavily on subjective peer assessments of

questionable validity and reliability. They also depend on data voluntarily self-reported by the universities themselves and therefore much more likely to be subject to institutional manipulation.³⁰ As governments seek to use market forces to coordinate and steer their university systems, they will need to define the essential performance information to be maintained and reported by universities.³¹ Public policy can thereby aid in the improvement of the reliability of information for student consumers, whether provided by the commercial sector or the not-for-profit sector.

Secondly, we have argued that information on subject fields and academic programs is of particular value to student consumers, even in North America where the structure of academic programs includes a strong emphasis on general education prior to the choice of a major field of study. While many North American colleges and universities collect information on the outputs of academic programs in their graduate placement offices, this information is rarely made public nor is it systematically used by the institutions themselves to improve the performance of academic programs. The types of program information that we believe should be publicly required of all institutions would include, at a minimum: entry standards for academic programs or subjects; program completion rates; the proportion of program graduates entering employment, professional training, and higher degrees; and the average starting salaries of graduates. Many colleges and universities in the US now collect such information as part of their placement services, but requiring this data as a condition for receiving government support or subsidy would not only make it more readily available for the developers of college guides and report cards, but would also likely foster more attention to these measures within universities.

Finally, the validity of league tables depends in part on their ability to provide measures that are closely approximate or clearly linked to the valued societal outcomes of higher education. This will require that university rankings provide more information on university processes and outputs, rather than the typical emphasis on university inputs. Indicators of the academic standard achieved by program graduates, the satisfaction of graduates with their academic programs, the academic legitimacy of academic programs, the effectiveness of institutional processes for assuring academic quality, and the extent to which the institution fosters behaviors known to be associated with effective student learning would all be examples of measures that could help strengthen the validity of university league tables for both student consumers and for university improvement. The development of such

data and information however would be expensive and is likely beyond the capacity of commercial publications. But several current national quality assurance policies produce information related to outputs and processes that have been or could be incorporated into university league tables as a means of making them more valid, relevant, and functional. Examples include:³²

- Output Indicators
 - National Course Assessments (Brazil)
 - Graduate Destination Survey and Course Experience Questionnaire (Australia)
- Process Indicators
 - Academic Audit (Australia, Hong Kong, New Zealand, Sweden, UK, US)
 - External Examiner System (Denmark, Hong Kong, Norway, UK)
 - National Survey of the Student Engagement (NSSE) US
 - Subject Assessments (Denmark, The Netherlands, Sweden, UK)

Without such government interventions to encourage the development of reliable and valid information about academic quality it is likely that commercially produced league tables will continue to under-serve the growing student consumer market and may also contribute as noted above to market failures in the performance of the overall higher education system.

Conclusion

As market competition in higher education becomes more common both within and across countries, governments are increasingly adopting strategies of information provision as a means of assuring academic quality. University league tables which compare the performance of different institutions have been advocated as a potentially efficient and effective means of providing needed information to student consumers as well as helping inform universities and policymakers on areas needing improvement.

Our review of the five leading commercial university league tables from Australia, Canada, the UK, and the US suggests that the definitions of academic quality used in these tables are converging. Whether this represents a global phenomenon needs to be more rigorously tested. As noted earlier, our sample may be biased and reflect cultural similarities characteristic of Anglophone countries. It would be valuable to

extend the analysis to league tables developed in Europe, Asia, and Latin America to see the extent to which a common construct of academic quality is becoming truly international. We also reviewed the effectiveness of these league tables applying a framework developed in a study of organizational report cards in various sectors. The five league tables varied in their validity, comprehensiveness, comprehensibility, relevance, and functionality. An apparently important contributor to the most effective university rankings is government policy. By specifying the performance indicators that will be publicly available and by subsidizing the development of measures of academic process and outputs, government can help improve the quality of information available to both student consumers and universities. This in turn will help assure the more effective functioning of competitive academic markets. In contrast, there is some evidence that more laissez faire policies on the availability of university performance information in North America may be contributing to less valid university league tables as well as to university behavior that does not benefit the public interest.

Acknowledgments

We wish to express our appreciation to Tom Karmel, Jamie Merisotis, David Weimer, and two anonymous reviewers for their valuable suggestions, but we remain solely responsible for the arguments advanced. We also wish to acknowledge the generous financial support of the Ford Foundation for the work leading to this paper.

Notes

1. A Summary Report from the meeting and selected papers were published in *Higher Education in Europe* 27(4), 2002.
2. Outside the US university rankings are often described as “league tables,” reflecting the published rankings used to place international football (i.e., soccer) teams in different leagues. The term league tables will be used in this paper as synonymous with commercially published university rankings.
3. Our criteria for selection of the league tables were commercial university rankings, which have been established long enough to provide some evidence and literature on their effectiveness and impacts. We have therefore focused on English-speaking countries, which may have biased the results in certain ways. We will address this issue in the Conclusion.
4. Teaching Quality Assessments (aka Subject Reviews) in the UK assessed the following aspects: Curriculum Design, Content, and Organization; Teaching,

Learning, and Assessment; Student Progression and Achievement; Student Support and Guidance; and Learning Resources, Quality Management, and Enhancement. However, as of 2001 systematic Teaching Quality Assessments have been discontinued by the QAA and some of the scores are now up to 10 years old. *The Guardian* has therefore, reduced the weight of TQA in its 2003 ranking.

5. Gormley and Weimer (1999) offer a sixth criterion, “reasonableness.” That is that a report card should be reasonable in the information demands placed upon the targeted industry and its organizations. Universities, for example, should have sufficient time for submitting necessary data and the costs of providing the requested information should not be excessive. As we will discuss, the information demands of governments with regard universities vary; in some countries all of the information included in university league tables is publicly provided. Therefore comparing the “reasonableness” of the various national report cards may be misleading and we have therefore chosen not to include this criterion in our analysis.
6. Note that the validity of measures is also dependent upon the reliability or accuracy of the relevant data or observations. There have been reports for example about the manipulation or misrepresentation of data used in university league tables. We will discuss this issue in the “Functionality” section.
7. Astin (1985) most clearly articulates this perspective on academic quality in his “talent development model.” Astin argues that the major purpose of a university is to develop the talents of its students to their maximum potential. This development is achieved by facilitating changes in students’ intellectual capacities and skills, values, attitudes, interests, habits, and mental health. Therefore, in Astin’s view, institutions that provide the largest amount of developmental benefits to students possess the highest academic quality.
8. Unlike Australia, Canada, and the US, the UK has conducted direct assessments of the quality of teaching through its TQA and these played a significant role in UK league tables. But research has suggested that TQA assessors have a positive research bias and university research ratings have been discovered to be a predictor of favorable TQAs (Drennan and Beck 2001).
9. Note, as reported above, the *GUG* assesses university “prestige” with objective measures of student demand, success in attracting research grants, and success in international ratings.
10. It is important to note that Astin’s research is based upon US data. Arguably the relationship between research and student learning may be more positive outside the US where the structure of first-level degree programs is different. However, the “American” approach to higher education – e.g. our hierarchical degree structure, the modular form of instruction, continuous assessment of students, de-emphasis on subject exams, and, most importantly, academic promotion and merit pay based primarily upon research and publication – is now being widely emulated throughout the world. It is possible that as these structural changes occur in higher education around the globe the negative relationship Astin detects between research and student learning may become more widely generalized.
11. For a comprehensive review of this economic research, see Winston and Zimmerman 2003.
12. This discussion is based on the analysis in a draft chapter on ‘Career and Economic Impacts of College’ kindly provided to us by Ernest Pascarella, which is included in the 2nd edition of Pascarella and Terenzini 2005.

13. The most recent performance indicators on graduate employment in the UK published by the HEFCE (2003) now provide benchmarks that adjust for a number of these variables.
14. Research on the Course Experience Questionnaire has provided support for its reliability and validity (Ramsden 1991; Wilson et al. 1997). But studies of the mandated surveys of graduates in Australia have found CEQ scores to be influenced by the response rate as well as by size of institution (Higher Education Division 2001).
15. The Australian Department of Education, Science and Training has done exemplary work in trying to adjust institutional indicators for student mix. Factors taken into account in adjusting indicators include gender, age, socio-economic status, field of study, indigenous status, level and mode of study. In 2001 the report took into account also the average Tertiary Entrance Ranking, which was found to explain a relatively large amount of variation in progress and attrition rate (Higher Education Division 2001). The *GUG* did not present these corrected scores.
16. A number of observers have recommended this type of categorization in UK league tables in order to acknowledge the observable differences in missions among universities (Drennan and Beck 2001; Provan and Abercromby 2000).
17. The major purpose of the NSSE is not to make university league tables more valid, but to provide information on educational processes that can help colleges and universities improve the quality of student learning. Perhaps more troubling is that few of the most highly ranked colleges and universities in the US even choose to participate in the survey.
18. While stressing the comprehensiveness of the TQA measure, we wish to acknowledge debates in the UK about the appropriateness of converting these complex assessments into single scores for league tables as well as challenges to the underlying reliability and validity of TQA data (see for example: Bowden 2000; Drennan and Beck 2001).
19. Entry qualifications to programs within the same university vary in the UK and Australia, where students are encouraged to choose a specific academic subject directly after secondary education. This is also true for some professional fields in the US and Canada such as architecture and engineering. But even comprehensive US universities, which traditionally encourage students to enroll in a "general college" program prior to selecting a subject or "major," are increasingly limiting later access to popular subjects such as journalism and business.
20. There is also reason to believe that in the US the relative influence of "general education" on student learning may be declining relative to the influence of the "major." Between 1914 and 1993 the average proportion of the graduation requirement composed of "general education" in the US declined from over half to less than a third and that composed of mandatory courses from a third to less than 7% (National Association of Scholars 1996). Research suggests that students' learning of academic content and their cognitive development is significantly associated with the "academic coherence" of the curriculum (Dill 1999). That is, student learning is affected by the pattern and sequence of the courses in which they enroll, by curricula requirements that integrate learning from separate courses, and by the frequency of communication and interaction among faculty members in the curriculum. Therefore, if an increasing proportion of the US undergraduate curriculum is becoming elective, it is likely that similar to their

peers in other countries, US university students derive the educational benefits of “academic coherence” primarily from their chosen academic subject or “major” field, if there.

21. Carricio et al. (1997) also demonstrate how university performance data used in league tables could be applied to construct a decision support system for university selection that responds to the differing preferences of university applicants.
22. Because it is difficult to isolate university responses to league tables in countries such as Australia and the UK, which have had a number of governmental initiatives in academic quality assurance, the following discussion will focus on the US where league tables have been a more significant instrument.
23. Both of these “corrections” were appropriate under *USNWR* definitions, but both also suggest how institutions are diverted into investing time and effort in activities unrelated to enhancing educational quality.
24. Because of concerns that institutions were manipulating their admissions processes in order to improve their rankings, *USNWR* dropped enrollment rate or “yield” (i.e. fraction of admitted students that accept an offer of admission) as a measure of academic selectivity in its 2003 issue on colleges and universities.
25. Ehrenberg (2002a; 2002b) also notes that the heavy weighting on institutional expenditures and faculty salaries in the *USNWR* formula provides an incentive to increase costs in higher education, since efforts to cut costs or increase productivity would lower an institution’s ranking.
26. Some have suggested UK universities have also “gamed” the RAE by recruiting productive researchers just prior to the rankings. Recent evaluations of the program have found little evidence to support this contention (Koelman and Venniker 2001).
27. We also recognize that universities are more sensitive to research doctoral peer ratings, because unlike reputation surveys at the first degree level the faculty members making research doctoral ratings can influence the quality of student input and the research grants awarded to the rated programs. Also, and unfortunately, it is likely that many university faculty members and administrators care more about the substantive quality of research doctoral programs than they do about the quality of first-level degree programs.
28. Hossler and Litten (1993) reviewed the overall provision of information on academic institutions in the US. They noted that virtually all of the college and university data used in commercial rankings were supplied by the institutions themselves and that no independent source of verification existed:

When colleges compete for students via the information they provide and the public must rely primarily upon this information, we find it intolerable that some form of audited and certified information, as precise and objective as our financial audits, is not available. (p. 78)

They called for standardized data gathering instruments and third party verification. Since they wrote a voluntary effort has emerged, the Common Data Set Initiative (see: www.commondataset.org), intended to provide a common set of standards and definitions for data used in university guidebooks and rankings, although no formal process of third party verification has been implemented.

29. We base our assessment of the Australian experience on personal reports from professional colleagues in that country. An analysis of the effects of these surveys is

- being undertaken as part of our PPAQ Research Program (see below). For evidence of the impacts of the TQA on universities in the UK, see Henkel (2000).
30. As previously noted there has been an effort in the US – led by the College Board, commercial publishers, and the US Association of Institutional Research – to develop a “Common Data Set” (CDS) that increases the reliability of the data used in university guidebooks and league tables. The data focuses primarily on inputs and financial aid and provides relatively little information on relevant processes and outputs as we have defined them. This is a voluntary effort, however, and lacks the force of law. It also does not seem to have reduced the amount of “gaming” of the league tables by various US institutions (see particularly Ehrenberg 2003).
 31. See for example the work of the Performance Indicators Steering Group (PISG) in the UK (HEFCE 1999), which defined information to be provided on the nature and performance of the higher education sector.
 32. The Research Program on Public Policy for Academic Quality (PPAQ) is conducting analyses of the development, implementation, and impact of these and other quality assurance instruments. For additional information, see the project website at: www.unc.edu/ppaq.

References

- Astin, A.W. (1985). *Achieving Educational Excellence: A Critical Assessment of Priorities and Practices in Higher Education*. San Francisco, London: Jossey-Bass Publishers.
- Astin, A.W. (1991). *Assessment for Excellence: The Philosophy and Practice of Assessment and Evaluation in Higher Education*. New York: American Council on Education/Macmillan.
- Astin, A.W. (1996). ‘Involvement in learning revisited’, *Journal of College Student Development* 40(5), 587–597.
- Becker, G.S. (1964). *Human Capital*. New York: Columbia University Press.
- Bowden, R. (2000). ‘Fantasy higher education: University and college league tables’, *Quality in Higher Education* 6(1), 41–60.
- Brennan, J. and Shah, T. (2000). *Managing Quality in Higher Education: An International Perspective on Institutional Assessment and Change*. Buckingham, UK: OECD, SRHE & Open University Press.
- Brewer, D., Gates, S.M. and Goldman, C.A. (2002). *In Pursuit of Prestige: Strategy and Competition in US Higher Education*. New Brunswick, NJ: Transaction Press.
- Carrico, C.S., Hogan, S.M., Dyson, R.G. and Athanassopoulos, A.D. (1997). ‘Data envelope analysis and university selection’, *The Journal of the Operational Research Society* 48(12), 1163–1177.
- Clarke, M. (2002). ‘Some guidelines for academic quality rankings’, *Higher Education in Europe* 27(4), 443–459.
- Connor, H., Burton, R., Pearson, R., Pollard, E. and Regan, J. (1999). *Making the Right Choice: How Students Choose Universities and Colleges*. London: Universities UK.
- Department for Education and Skills (DfES) (2003). *The Future of Higher Education*. London: HMSO.

- Dill, D.D. (1999). 'Student learning and academic choice: The rule of coherence, in Brennan, J., Fedrowitz, J., Huber, M. and Shah, T. (eds.), *What Kind of University? International Perspectives on Knowledge, Participation and Governance*. Buckingham: Open University Press and Society for Research into Higher Education, , pp 56–70.
- Dill, D.D. (2000). 'Designing academic audit: Lessons learned in Europe and Asia', *Quality in Higher Education* 6(3), 187–207.
- Dill, D.D. (2003). 'Allowing the market to rule: The case of the United States', *Higher Education Quarterly* 57(2), 136–157.
- Drennan, L.T. and Beck, M. (2001). 'Teaching quality performance indicators – Key influences on the UK universities' scores', *Quality Assurance in Education* 9(2), 92–102.
- Eccles, C. (2002). 'The use of university rankings in the United Kingdom', *Higher Education in Europe* 27(4), 423–432.
- Ehrenberg, R.G. (2002a). 'Reaching for the brass ring: The *U. S. News & World Report* rankings and competition', *The Review of Higher Education* 26(2), 145–162.
- Ehrenberg, R.G. (2002b). *Tuition Rising: Why College Costs so Much*. Cambridge: Harvard University Press.
- Ehrenberg, R.G. (2003). 'Method or madness?: Inside the *USNWR* college rankings', Paper presented at the Wisconsin Center for the Advancement of Postsecondary Education Forum on *The Abuse of College Rankings*, Madison, Wisconsin, 20–21 November.
- Ehrenberg, R.G. and Hurst, P.J. (1996). 'A hedonic model', *Change* 28(3), 46–51.
- Gormley, W.T.Jr. and Weimer, D.L. (1999). *Organizational Report Cards*. Cambridge, Mass.: Harvard University Press.
- Graham, A. and Thompson, N. (2001). 'Broken ranks: *US News*' college rankings measure everything but what matters. And most universities don't seem to mind', *Washington Monthly* 33(9), 9–14.
- Grunig, S.G. (1997). 'Research, reputation, and resources: The effect of research activity on perceptions of undergraduate education and institutional resource acquisition', *Journal of Higher Education* 68(1), 17–52.
- Higher Education Division, Australia (2001). *Characteristics and Performance Indicators of Australian Higher Education Institutions*. Occasional Paper Series 01-B, Department of Education, Science and Training: <http://www.dest.gov.au/highered/statistics/characteristics/characteristics00.pdf>
- Higher Education Funding Council for England (HEFCE) (1999). *Performance Indicators in Higher Education: First Report of the Performance Indicators Steering Group (PISG)*. Report 99/11, Higher Education Funding Council for England: <http://www.hefce.ac.uk/pubs/hefce/1999/99%5F11.htm>
- Higher Education Funding Council for England (HEFCE) (2003). *Performance Indicators in Higher Education: 2000–2001 and 2001–2002*. Report 2003/59, Higher Education Funding Council for England: <http://www.hefce.ac.uk/learning/perfind/2003/default.asp>
- Henkel, M. (2000). *Academic Identities and Policy Change in Higher Education*. London: Jessica Kingsley.
- Hossler, D., Braxton, J. and Coopersmith, G. (1989). 'Understanding student college choice', in Smart, J.C. (ed.), *Higher Education: Handbook of Theory and Research, Vol.V*. New York: Agathon Press, pp. 231–288.

- Hossler, D. and Litten, L.H. (1993). *Mapping the Higher Education Landscape*. New York: College Entrance Examination Board.
- James, R., Baldwin, G. and McInnis, C. (1999). *Which University? The Factors Influencing the Choices of Prospective Undergraduates*. Canberra: Australian Government Publishing Service.
- Koelman, J. and Venniker, R. (2001). 'Public funding of academic research: The research assessment exercise in the UK', in CPB and CHEPS (eds.), *Higher Education Reform: Getting the Incentives Right*. Den Haag, The Netherlands: Sdu Uitgevers, pp. 101–117.
- Kuh, G. (2003). 'What we're learning about student engagement from NSSE', *Change* 35(2), 24–32.
- Larsen, K., Martin, J.P. and Morris, R. (2002). 'Trade in educational services: Trends and emerging issues'. Paper Presented at the OECD Forum on Trade In Educational Services, May 23–24, Washington DC.
- McDonough, P.M., Antonio, A.L. and Perez, L.X. (1998). 'College rankings: democratized college knowledge for whom?', *Research in Higher Education* 39(5), 513–537.
- Manski, C.F. and Wise, D.A. (1983). *College choice in America*. Cambridge, MA: Harvard University Press.
- Merisotis, J.P. (2002). 'Summary report of the invitational roundtable on statistical indicators for the quality assessment of higher/tertiary education institutions: Ranking and league table methodologies', *Higher Education in Europe* 27(4), 475–480.
- Monks, J and Ehrenberg, R.G. (1999). 'U.S. News & World Report's college rankings: Why they do matter', *Change* 31(6), 42–51.
- Moogan, Y.J., Baron, S. and Harris, K. (1999) 'Decision-making behaviour of potential higher education students', *Higher Education Quarterly* 53(3), 211–228.
- Morrison, H.G., Magennis, S.P. and Carey, L.J. (1995). 'Performance indicators and league tables: A call for standards', *Higher Education Quarterly* 49(2), 128–145.
- National Association of Scholars (1996). *The Dissolution of General Education: 1914–1993*. Princeton: National Association of Scholars.
- National Opinion Research Center (1997). A Review of the Methodology for the *US News and World Report's* Rankings of Undergraduate Colleges and Universities. *Washington Monthly*: <http://www.washingtonmonthly.com/features/2000/norc.html>
- Page, S. (1999). 'Rankings of Canadian universities and help to students', *Guidance & Counseling* 14(3), 11–14.
- Pascarella, E.T. (2001). 'Identifying excellence in undergraduate education: Are we even close?', *Change* 33(3), 19–23.
- Pascarella, E.T. and Terenzini, P.T. (1991). *How College Affects Students: Findings and Insights from Twenty Years of Research*. San Francisco: Jossey-Bass.
- Pascarella, E.T. and Terenzini, P.T. (1998). 'Studying college students in the 21st century: Meeting new challenges', *The Review of Higher Education* 21(2), 151–165.
- Paulsen, M.B. (1990). *College Choice: Understanding Student Enrollment Behavior*. ASHE-ERIC Higher Education Report No 6. Washington, DC: The George Washington University, School of Education and Human Development.
- Provan, D. and Abercromby, K. (2000). University League Tables and Rankings: A Critical Analysis. Commonwealth Higher Education Management Service (CHEMS) Paper No. 30: <http://www.acu.ac.uk/chems/onlinepublications/976798333.pdf>.

- Ramsden, P. (1991). 'A Performance indicator of teaching quality in higher education: The course experience questionnaire', *Studies in Higher Education* 16(2), 129–149.
- Rosovsky, H. and Hartley, M. (2002). *Evaluation and the Academy: Are We Doing the Right Thing?* Cambridge, Mass.: American Academy of Arts and Sciences.
- Smith, J., McKnight, A. and Naylor, R. (2000). 'Graduate employability: Policy and performance in higher education in the UK', *The Economic Journal* 110(June), 382–411.
- Stecklow, S. (1995). 'Colleges inflate SAT's and graduation rates in popular guide-books', *Wall Street Journal* 5 April.
- Terenzini, P.T. and Pascarella, E.T. (1994). 'Living with myths: Undergraduate education in America', *Change* 26(1), 28–32.
- Thompson, N. (2000). 'Playing with numbers: How *US News* mismeasures higher education and what we can do about it', *Washington Monthly* 32(9), 16–23.
- Tight, M. (2000). 'Do league tables contribute to the development of a quality culture? Football and higher education compared', *Higher Education Quarterly* 54(1), 22–42.
- Trow, M. (1999). 'Biology at Berkeley: A case study of reorganization and its costs and benefits', Center for Studies in Higher Education: Research and Occasional Papers Series, CSHE.1.99: <http://ibb.berkeley.edu/cshe/>
- Trow, M.A. (1983). 'Organizing the biological sciences at Berkeley', *Change* 15(8), 28, 44–53.
- Wilson, K., Lizzio, A. and Ramsden, P. (1997). 'The development, validation and application of the course experience questionnaire', *Studies in Higher Education* 22(1), 3–25.
- Winston, G.C. and Zimmerman, D.J. (2003). 'Peer effects in higher education', NBER Working Paper 9501: <http://www.nber.org/papers/w9501>.
- Yorke, M. (1997). 'A good league table guide?', *Quality Assurance in Education* 5(2), 61–72.
- Yorke, M. (1998). 'The Times "League Table" of universities, 1997: A statistical appraisal', *Quality in Education* 6(1), 58–60.
- Yorke, M., Barnett, G. Bridges, P., Evanson, P., Haines, C., Jenkins, D., Knight, P., Scurry, D., Stowell, M. and Woolf, H. (2002). 'Does grading method influence honours degree classification?', *Assessment & Evaluation in Higher Education* 27(3), 269–279.
- Zemsky, R. and Oedel, P. (1983). *The Structure of College Choice*. New York: College Entrance Examination Board.