The Canadian Journal of Higher Education
La revue canadienne d'enseignement supérieur
Volume XXXV, No. 3, 2005 pages 27 - 55

"World Class" or The Curse of Comparison?

DANIEL W. LANG
University of Toronto

ABSTRACT

Can all the universities that claim to be "world-class" actually live up to the claim? If they could be, would that be desirable public policy? It could be that there are so many different meanings of "world-class" that the term in practical effect is an oxymoron: the definition of "world" is determined locally when conceptually it should be defined internationally. This paper discusses different kinds of institutional quality, how quality is formed and how it can be measured, particularly by comparison. It also discusses the subtle but fundamental differences between quality and reputation. The paper concludes with the suggestion that world-class comparisons of research quality and productivity are possible, but that any broader application to the "world-class" quality of universities will be at best futile and at worst misleading.

RESUMÉ

Toutes les universités qui se disent de « classe mondiale » peuvent-elles vraiment se montrer à la hauteur de cette revendication ? Si elles le pouvaient, est-ce que ce serait dans l'intérêt public ? En fait, l'expression « classe mondiale » peut prendre tant de sens qu'il s'agit dans la pratique d'un oxymore : la définition de « mondiale » est établie à l'échelle locale alors qu'en fait elle devrait l'être au plan conceptuel à l'échelle internationale.

Cet article traite des différentes sortes de qualités caractérisant les établissements, de la manière dont ces qualités en viennent à exister et de la façon de les évaluer, particulièrement en faisant des comparaisons. Il analyse également les différences subtiles, mais fondamentales, entre qualité et réputation. En conclusion, l'article suggère qu'il est possible de comparer le niveau de la qualité et de l'efficience des recherches des universités à l'échelle mondiale, mais qu'une application plus vaste de leur qualité de « classe mondiale » serait futile, voire trompeuse.

INTRODUCTION

At the end of each installment of his popular *Prairie Home Companion* radio stories about the fictional American town of Lake Wobegon, Garrison Keillor describes the town as being a place where "all the children are above average." If one were to take into account all the claims made by universities about being "world-class" the impression would be similar: they are all world-class. But, just as not all children can be above average, not all universities can be world-class. Even if they could be, perhaps they should not be.

This is more than a matter of puckish humour. It begs some important questions. What does "world-class" mean in higher education? Will a headlong rush to be world-class – whatever it might be mean – defeat diversity and compromise accessibility and possibly even quality? What are the costs and benefits of what amounts to "look alike competition" which instead of stimulating innovation encourages conformity? Who should be the arbiter of eligibility for membership in the club of world-class universities? Although the question may seen counter-intuitive, is it possible, given the costs of being world-class, that world-class status is not infinitely valuable, affordable, or manageable? Can there be too many world-class universities just as there can be too few?

These are difficult questions. They become even more difficult when one asks whether world-class is a system concept or an institutional concept. At first that may seem to be a question with only one answer: world-class is only an institutional concept. But a pause for reflection will reveal that the question is not that simple. For example, although public universities have different degrees of autonomy, they all depend

on governments for financial support, they have mandates that, at least to some extent, are set by the state, and they are parts of larger systems of higher education, some of which are highly centralized. Unless one were to disqualify all public universities from the world-class rank, being world-class must have a system dimension. Because in some jurisdictions there are large public subsidies to otherwise private universities, as in the case of many student financial aid programs and research funding agencies, even some private universities are defined by public policy. If being world-class is for some universities partly an expression of the performance of a system of higher education, diversity, logically, must be linked to quality.

Diversity among institutions or, at least, institutional types, is a policy objective that most systems of higher education pursue. At the same time those systems are also concerned about equity of access and the quality of educational opportunity. In highly differentiated systems, world-class is a deliberate part of diversity, as the existence of "flagship" campuses in jurisdictions like California in the United States, of the *grandes ecoles* in France, and the 100 "world standard" universities in China (Lang and Zha, 2004), demonstrate. In another time, "flagship" and "world-class" might very well have meant the same thing.

Also in another time "reputation" and "quality" might have been taken to be synonyms. Research into "college choice" – that is, the factors that influence students' decisions about the programs and institutions that they will apply to and attend – typically presents reputation and quality as separate factors in surveys and interviews of student choices (McDonogh 1998, Clarke 2002, Lang and Lang 2002). In the 1970s and 1980s, quality, expressed in different ways, usually ranked at the top as the most influential factor in students' choices. Reputation usually ranked around the bottom of the first quartile, and sometimes lower. In the 1990's the pattern began to change; reputation rose and in some cases displaced quality as the most influential factor. This perhaps should not be surprising. We know that universities regularly engage in marketing and recruitment, and that "branding" is an institutional strategy for the purposes of attracting students and research funding. These strategies are not necessarily the desperate acts of universities that are far from the mark in terms of being world class. One has to look only as far as M.I.T. and Columbia in the

United States, *Universitas 21* internationally, or Cambridge in England to find top-ranked universities promoting their reputations as brand names for the purposes of marketing.

However the differences and similarities between reputation and quality are resolved, another disparity will remain. Are reputation and quality about education or research? The reward system for university faculty favours research. To the extent that journalistic rankings of universities are rewards to institutions, they too tend to favour research. In the case of both types of reward, there is no policy or deliberate intention to be lop-sided in favour of research. Nevertheless that is how the arithmetic of many league tables works, perhaps for no other reason than the broad availability of comparable data and the portability of measures of research across national boundaries (Bok 2003, Altbach 2004). A recent ranking of the world's top 500 universities is based entirely on measures of research performance (Liu 2004).

From the start, then, there is no logical reason to expect that there can be a single definition of "world-class." Instead, the question has to be about the different parts or facets of what being world-class means. These components include quality, reputation, and context – that is, world-class as a concept within systems of higher education and as a concept that applied to universities independently. Finally, there is the implication that being world-class represents a trade-off with diversity, or at least a different view of public policies that promote diversity in higher education.

Quality

It would be easy to say, simply, that there is no agreement about what quality means in higher education, or that, like art, quality is in the eye of the beholder. As a matter of fact, it's true: there is no firm consensus internationally about what world-class quality is (Frazer 1994). "Internationally" is a key word in this context. It is not surprising to find that individual jurisdictions have devised definitions and measurements of quality that suit their own institutions. But those definitions and measurements do not necessarily lend themselves to international or worldwide application. There are plausible reasons that explain why definitions of quality are often not internationally portable.

First among these is audience: different stakeholders have different expectations about quality. Governments, as the principal financiers and regulators of higher education, typically take what in the end amounts to a cost-benefit view of quality in universities. This view also forces an ineluctable tendency to evaluate the performance of universities along with the performances of schools, hospitals, and social welfare. In other words, the basic idea of quality becomes a matter of proximate and optimal value instead of absolute value. It also becomes essentially comparative and relative.

Although governments purport to represent the interests of individual citizens, and therefore behave as if they held the proxies of all stakeholders in defining and appraising world-class quality, human capital theory explains why that representation is a political conceit. Moreover and more to the point, it is ironic. An argument is often advanced that markets, competition, and autonomy are the ingredients for world-class performance. That may or may not be so, but for the question at hand, there is enough confirming scholarship, both anecdotal (Altbach 2004, MacTaggart 1998, El-Khawas and Massy 1996) and empirical (McLendon 2003, Clark 1998, Volkwein 1986) to suppose at least that it is.

A fundamental precept in the economics of education is that investments, public and private, in higher education generate economic returns, also both public and private. That is what human capital theory is about. Governments act on that theory when they provide public subsidies to colleges and universities, and when they regulate tuition fees as private subsidies. Independent universities – either for-profit or not-for-profit – set tuition fees on the assumption that students as buyers will perceive that at least some economic value will arise from their investments. There is much wariness about higher education in the marketplace, but as long as there are private investments, one must assume that students, and in many cases their families, may have an interest in quality and world-class standing that is different from the interest of government. For example, because governments are the major financiers of university research as well as universities generally, some of them have relatively little interest in separating education and research for the purpose of

defining quality. Some of them may in fact expect universities to rob the instructional Peter in order to pay the researcher Paul, which they often do. (Ehrenberg 2003).

Students on the other hand may have a very strong interest in making the separation. As Adam Smith said in *The Wealth of Nations*: "The discipline of colleges and universities is in general contrived, not for the benefit of the students, but for the interest, or more properly speaking, for the ease of the masters." (Smith 1937, III, II, 3) Smith would have good reason to say the same thing were he magically to be transported to the 21st century. As the cost of tuition rises, often on the belief that investments in higher education will generate returns far in excess of costs, students want to connect their investments to the quality of instruction. Governments and universities often promote higher education with explicit promises about returns on investment (Wolf 2002).

In some jurisdictions – Britain and Australia, for example – governments are beginning to finance research and education separately. Within the context of this discussion, the most significant aspects of these initiatives is that research is being assessed in terms of quality while undergraduate education is being assessed in terms of productivity, efficiency, and added value. Quality, as an expression of world-class status, is becoming increasingly more about research than education (Altbach 2004).

Human capital theory also explains why employers, first, have an interest in understanding what quality means in higher education and, second, have an interest that may be different from the interests of government and students. The general public may have an interest closely related to the interest of employers in the sense that definitions of quality – especially those that arise from accreditation – can serve as a sort of consumer protection or assurance about the qualifications of graduates.

There is a final audience: universities themselves, particularly the professoriate, are concerned about definitions of quality. Indeed, they are the principal source of such definitions. Within the context of world-class quality, it is important to recall that the university as an instrument of the state is a relatively new concept. Scholarship, especially research, has a long and laudable history of portability among nations (Winchester 1986). Perhaps the tendency to attach the quality of research (as opposed

to the quality of instruction) to world-class status should not be surprising. It may be a very old idea that is being recycled as it is in new research assessment procedures that have been recently introduced in Britain, the United States, Hong Kong, and Australia (El-Khawas and Massy 1996). These procedures rely heavily on the judgment of disciplinary peers as opposed, significantly, to institutional or governmental agents. Moreover, these peers are often international. For example, the definition of what is "world-class" in university chemistry may be being determined by chemists, some of who may be employed by private R&D firms, without regard to organizational or jurisdictional boundaries.

Because each of these audiences is different and has a different reason for being interested in quality in higher education, that there are different definitions of quality should not be surprising. Indeed, it is ineluctable. For a university or a government to pursue, as a matter of policy, world-class status, that university or government must explicitly determine what its peer audiences are. It must also understand the factors and forces that cause colleges and universities to change, in this case in the direction of attaining world-class status.

Reputation

Aren't quality and reputation the same? Perhaps they are not the same at all or, at least, not as alike as they might at first seem. There are two sets of facts that elucidate what otherwise might seem to counter intuition.

First, in a study conducted at the University of Toronto (Lang and Lang 2002) of elite students who were applying to American and Canadian colleges and universities, over two-thirds of the students said that they took the difficulty of being admitted as a surrogate for institutional quality. That assumption was especially pronounced – over 90 per cent – among the students who were applying to an American college or university. They indicated further that the admissions selection process was evidence of an intense competition for admission. An offer of admission to a highly selective university is coming to be viewed as a trophy with an inherent value of its own. One student who was interviewed made the point well: "They have to do that because they have so many applicants with high grades and high SATs that they have to find other ways of picking students."

That student was closer to the point than she might have realized. As the level of competition for admission intensifies from what the Carnegie Council called "admissibility" to "selection" the process of selection shifts away from academic competence to personal characteristics and traits that "commend themselves for consideration" (Carnegie Council on Policy Studies in Higher Education, 1977).

Because of the rising cost of tuition, even at public colleges and universities, and because of reports from guidance counselors that good students sometimes make bad college choices, the College Choice Project at the University of Toronto also sought to determine the extent to which applicants drew distinctions between quality and reputation as factors in their choices of college and university. They did. Just over 70 per cent indicated that in terms of future job opportunities and of subsequent admission to graduate schools and professional schools, reputation might be more valuable than quality. In the survey component, where quality and reputation might have been regarded as different versions of the same query, a clear difference was perceived as each factor received different scores. Quality ranked higher – just under 90 per of all respondents "strongly agreed" – than reputation as just under 80 per cent "strongly agreed" that it was a factor in their college choice.

Compared to other post-admission surveys (Astin, 1993; Acumen, 1998, 1999) the relatively high ranking of reputation is significant. Parents, in that part of the College Choice project, ranked reputation higher than quality. This was true of virtually all parents of students who were applying to American colleges and universities. These results are similar to those found in an American study of students in Ivy League colleges and universities who were asked to rank the factors that influenced their college choices. "Prestige and name recognition" ranked first, slightly but nevertheless significantly ahead of "academic program" (Greene 1998). In the College Choice Project, this was the area in which the views of guidance counselors most differed from the views of students. Counselors tried to discourage students from selecting colleges and universities on the basis of their reputations instead of the quality and suitability of their programs; the majority of students discounted that advice.

Second, the arithmetic mechanics of major surveys of nominal quality do not regard reputation as a synonym for quality. The major American survey is conducted annually by *U.S. News & World Report.* The counterpart of the *U.S. News & World Report* survey in Canada is conducted by *Maclean's* magazine. In both of these surveys reputation is one factor among many, and in neither case is a major factor. Reputation comprises no more than 20 per cent of the *Maclean's* survey and 25 per cent of the *U.S. News & World Report* survey (Morse and Flanagan 2002). Thus while there is some overlap between quality and reputation it is relatively minor.

The sources of the information that *U.S. News & World Report* and *Maclean's* collect about reputation also indicate a difference between quality and reputation. The *Maclean's* survey bases its measure of reputation on rates of financial support from alumni, and on a poll of corporate chief executives, secondary school guidance counselors, and university administrators. Of these three, the third group is the only one that could provide credible information about quality. The other two groups may know something about reputation, but are unlikely to be reliable judges of quality. *U.S. News & World Report* surveys only college and university presidents and deans of admission. Neither survey polls professors and researchers (unless they hold administrative positions in addition to their academic positions).

There is some reason to believe that even the university administrators' perceptions of quality may be in the first instance reputational when applied to institutions other than the ones by which they are employed and therefore directly familiar (Webster 1992). What this boils down to is relatively simple: the statistical structure of commercial rankings of universities inherently separates quality and reputation.

In 2004 Gary Pike conducted a comparative analysis of the *U.S. News & World Report* survey and the [American] National Survey of Student Engagement (NSSE) which measured the involvement of students in activities that contribute to academic performance. Pike's first and strongest conclusion was that "the quality of a student's education is not synonymous with the resources and reputation of an institution." (p. 204) This conclusion is the quantitative counterpart of the qualitative reports

from the College Choice Project about the efforts of guidance counselors to persuade students define quality apart from reputation (Lang and Lang 2002)

Because rankings of colleges and universities are usually constructed by for-profit press media, the fact that they are commercially successful indicates that there is a market for information about reputation. The results of the College Choice Project at the University of Toronto explain the origin of that market: students want to know about reputation and they do not necessarily equate quality with reputation Does this make sense? It does to economists. (Lang and Lang 2002)

Human Capital Theory and Signaling

Benjamin Franklin said "If a man empties his purse into his head, no man can take it away from him. An investment in education pays the best dividends." Poor Richard, Franklin's *nom de plume*, would have called that common sense. Today economists call it the theory of human capital. It is a powerful theory that simultaneously explains large-scale public subsidies to colleges and universities and high tuition fees at private colleges and universities.

The idea is not complex: human intelligence and skill can produce wealth just as land and machines – that is, conventional capital – produce wealth. Education adds to the value of human capital and, in turn, generates public monetary returns in the form of higher productivity and economic growth. It also generates private monetary returns in the form of higher earnings.

There is a causal relationship between the quality of education and human capital theory: greater quality logically ought to produce an expansion of human capital and, in turn, higher returns, both public and private. Why then would there be a differentiation between quality and reputation? The answer is another economic theory usually called "signaling" but sometimes also called "screening." Here the idea is still based on human capital, but postulates that many of the human qualities that create wealth – intelligence, industry, initiative, and innovation, for example – are innate, and that formal education adds little to their productive value (Berg 1971). According to signaling or screening theory,

what colleges and universities do is rank or sort students in terms of their innate qualities (Rubenstein 1998) thus indicating to employers which students present the most human capital, much of which they possessed prior to becoming post-secondary students. In other words and with regard to a previous point, offers of admission can alone really be economically valuable trophies.

The screening function does not add to human capital. It is not an expression of educational or institutional quality. It is an expression of student quality that uses the language of institutional reputation. This is why, if we return to the arithmetic of league rankings, the *U.S. News & World Report* and *Maclean's* surveys assign high weights to the high school grades and standardized test scores of entering students. This is why students ignore the advice of guidance counselors to choose universities on the basis of quality instead of reputation. This is why alumni of universities are among the most frequent purchasers of the *U.S. News & World Report* and *Maclean's* surveys. This is also why *Money*, which is another magazine that ranks American colleges and universities, openly says that its survey is not about quality; it is about value for money (Webster 1992). The *U.S. News & World Report* and *Maclean's* surveys both now include sub-tables that purport to measure added value and "best buys."

The signaling function presents a number of contradictions in the ranking of colleges and universities. First, as just explained, the market signals for higher education are in the first instance about reputation, and only coincidentally about quality. But through a combination of funding policies based on human capital theory and neo-liberal global economics, higher education has been marketized, particularly among research-intensive universities (Slaughter and Leslie 1997, Ehrenberg 2003). This leads to the second contradiction: for the higher education market to work, students and governments as purchasers need to know more than they do about the quality of the colleges and universities in which they invest. Publications like *U.S. News & World Report*, *Money*, and *Maclean's*, whatever their failings, are *de facto* measures to fill a gap in information that the market needs to function properly (Webster 1992, McDonogh 1997). In some jurisdictions –Ontario, Canada, for example – governments have deployed "key performance indicators" in an attempt

to correct market imperfections by injecting additional and standardized information for students as consumers.

This gap is a serious problem for any attempts to appraise the quality of a college or university on the basis of student selectivity (which in some cases accounts for as much as 40 per cent of non-reputational rankings). When Michael Spence was awarded the Nobel Prize in 2001, he was asked by a journalist "whether it was true that you could be awarded the Nobel Prize in Economics for simply noticing that there are markets in which certain participants don't know certain things that others in the market do know?" (Spence 2001). The answer, of course, was yes: the degree of aysmmetry, as Spence used the term, was, if not simple, surprising. In economic terms, the market for higher education is highly asymmetrical. This is exemplified by the research on college choice that reveals the erroneous but forceful perceptions that applicants have about the selection process (Lang and Lang 2002).

Governments also are buyers in the higher education market, or perhaps in two higher education markets. To the extent that governments provide subsidies to colleges and universities in order to increase productivity and stimulate economic growth, they are making investments or purchases from which they expect certain returns. In these cases the purchase is of education, or in economic terms, increased human capital. Governments thus are interested in quality even if students and their parents are interested in reputation.

In other cases, the government is the principal in a principal-agent relationship for the acquisition of research. It is a particularly difficult principal-agent relationship because, almost by definition, the principal as buyer cannot know at the time of purchase the outcome of the research that the agent – a university researcher – will undertake. In other words, the market for university research is ineluctably and inherently asymmetrical. Discussing university research in these terms may seem crass, but it explains why reputation is often a surrogate for quality. For example, the ranking of the world's top 100 universities constructed by the Institute for Higher Education at Shanghai Jiao Tong University (Liu 2004) is based entirely on measures of research performance, and every measure is essentially an historical track record. In some cases the record extends

as far back as 1911 and 1981. In the higher education research market, in terms of the information available to the funding agency as purchaser, virtually the only information available is about reputation.

This poses a third contradiction. Let us assume that reputation is an acceptable surrogate for quality, and that the various institutional measurements and rankings based on reputation are accurate. Of course these assumptions are questionable but that is not the point. The point, instead, is more profound and potentially troubling. If rankings of world-class standing constructed in this way are taken seriously, for example, by allocating funding according to rank or by setting tuition fees on the basis of rank, they may impede the creation of knowledge by reinforcing obsolete reputational paradigms and homogenizing scholarship. In that case, winning the race to be number one may in the end be an accomplishment of Pyrrhic proportion.

Quality by Comparison

However world class quality is defined, it is inevitably a comparative concept. Colleges and universities throughout much of the world are regularly compared to one another. Sometimes they do this by their own choice, usually either to benchmark their costs and performance or to determine their competitive market positions. At other times they are compared and ranked by the press, a practice that most colleges and universities on the one hand, decry, and, on the other hand, cannot resist (Just 2002).

At yet other times, usually in the name of accountability, governments whose systems of higher education are centralized compare colleges and universities, at least those within their own systems. Sometimes the systems themselves are compared, in which case quality often is not the primary index. Often the degree of diversity within each system becomes a point of comparison, usually with respect to questions about how differentiation among institutions might be measured and promoted, and how distinctive institutional missions and roles might be recognized within each system.

Measuring quality and diversity is not easy. This is not only a matter of objective quantification and comparison. It also has a lot to do with how universities change in order to become world-class or more diverse. What

is the fundamental cause of change in higher education? What events and conditions might cause a university to become world class? This is not a matter of simple instrumentalism. For example, as Pike's research (2004) and Volkwein's earlier research (Volkwein 1989) demonstrate, infusions of resources do not necessarily improve or signify quality in higher education. In turn, neither is it necessarily possible for a university to spend its way to world-class status.

Robert Birnbaum, who has written extensively about diversity in higher education identified at least six different kinds of diversity and two different paradigms – "natural selection" and "resource dependence" (Birnbaum 1983). He and others further observed that none of the conventional, broadly applied institutional classification schemes – including some of those that purport to identify "world-class" – satisfactorily accounts for all institutional characteristics, including quality (Birnbaum 1983).

There are other paradigms. Joseph Ben-David (1972) argued that universities change is response to competition, and that competition is greatest when colleges and universities are relatively independent. Altbach (2004), Clark (1998), and MacTaggart (1998) later advanced similar cases for institutional autonomy *cum* competition. This would imply a paradigm rooted in organizational behaviour and system structure. From this follows an intriguing paradox: as governments pursue quality through the construction of more highly regulated and planned systems of higher education they may in practical fact be creating environments that discourage quality by over-generalizing about it.

This in turn suggests another question: Is it quality and diversity in terms of world class that should be measured or is it the conditions that engender world class standing, in this case the level of regulation, including accreditation, which should be measured? Since regulation is an almost exclusively system concept, and since institutional change and differentiation is a continuous process (Blau 1994), comparisons based on individual institutions, regardless of how they are classified, might be a step away from the real issue. Assuming that there is an agreement on what being world-class means, and that there is a reliable methodology for determining when a university has reached world class status, an important question remains about what that university must do to retain its status.

Peter Blau, in *The Organization of Academic Work*, a title that itself suggests a theory about the foundations of world-class standing, advanced a paradigm based on social forces, institutional size and the proportionate scale of administration. According to Blau (1994), these factors operate in more or less the same way regardless of institutional type. An implication is that the classification of institutions by group is not a reliable measure of world class standing.

"World-class" thus can have a system meaning, an institutional meaning, and a disciplinary meaning. Also, world-class standing is, in system terms, a form of diversity to which only some institutions either can or should aspire (Altbach 2004).

Quality, Reputation, and the Selection of Peers

Just as guilt by association is a specious proposition, so is being world-class by association. Defining world class standing by association is reliable only when there is an objective and systematic means of determining peers for the purpose of comparison and ranking. Peer selection as a policy issue began to grow in importance as interest in accountability and performance indicators grew, and as colleges and universities came under greater pressure to perform efficiently. In order to make informed decisions about strategy and resource allocations individual institutions might quite legitimately wish to construct comparisons with other institutions for the purposes of benchmarking. Benchmarking is not necessarily about quality. More often it is about the efficient use of resources, usually in monetary terms, but not always. For example, the utilization of space is often benchmarked.

There are many different indicators of performance, and almost as many debates about their reliability, relevance, and fundamental purposes. Nevertheless, most public systems of higher education are committed to them (Lang 2005). As well and more to the point, any attempt to assign world-class status on the basis of performance indicators is inherently comparative.

Comparisons made *ad hoc*, either because data are readily available or because comparisons with certain other institutions produce intuitively desirable results, are inherently unreliable. Convenience and politically useful results should not form the basis of peer selection. Nor can the

quality and performance of individual colleges and universities or systems of higher education be effectively appraised by anecdote. Yet, in the absence of systematic means of determining peers, that is an entirely possible and unfortunately misleading result.

There are many legitimate criticisms of schemes that rank colleges and universities. At the same time, as Michael Spence (2001) might say, there is a need for more information about the performance of colleges and universities so that the "buyers" in the higher educational marketplace know as much as the "sellers." Most critics of ranking focus on its statistical methodology and the legitimacy of the various factors that are used to express institutional performance. However, even if those deficiencies were corrected, a major problem would remain. Unless one presumes that every university has the same mission, and that every university has the same access to resources, inter-institutional comparisons – especially international comparisons – can fail. There are two methodologies in the foundation on which ranking schemes should be built. The first comprises the weights, indices, and definitions that measure and express institutional performance. The second comprises the ways and means of selecting peers for the purpose of comparison. The first methodology has received a lot of scholarly attention. The second has not (Lang 2000).

Peer selection is as much an art as a science, and fundamentally involves professional judgment. The ultimate objective of any methodology for determining peers for comparison should be to ensure that the institutions are sufficiently similar for comparisons to make sense. Institutions have different roles, some deliberately set as mission statements while other roles are the products of history; others still are the unfortunate consequence of institutional drift. Institutions are different in terms of size and location. They are different in terms of organizational complexity, which is not necessarily determined by size.

These differences are more than technicalities. If rankings of world class standing were more complex and, particularly, balanced instruction and research, graduate and undergraduate education, and reputation and quality, what otherwise might seem to be minor differences among institutions would indeed be minor. The problem is that no national or international ranking scheme does that. The *U.S. News & World Report*

and *Maclean's* surveys try to solve part of the problem by placing different types of institutions in different categories. Not all of those types are internationally portable. A notable example are the elite American liberal arts colleges like Amherst and Swarthmore that, as an institutional type, do not exist outside the United States but whose quality and reputation for undergraduate education often exceed those of their of Ivy League competitors. In terms of resources on a per student basis – which is frequently taken as a measure of quality – they rank even further ahead. What this example suggests is that the word "class" in world class should have two meanings, with the second meaning denoting a generic type of institution. This example also suggests, as recent research indicates (Marsh and Hattie 2002), that the quality of instruction and the quality of research are not so causally linked that any ranking of world-class status must incorporate both.

Furthermore, Blau's paradigm of institutional change, like Diana Crane's (1972) work on *Invisible Colleges*, suggests that, in terms of the institutional behaviours that lead to world class-ness, fundamental academic change occurs in the first instance within disciplines and fields of knowledge. It is only secondarily organizational or structural. The logic of this paradigm is that while there might be comparatively defined world-class departments of, for example, chemistry, comparative definitions of world-class institutions are problematic. This paradigm may also explain the trend to express quality as a measure of research performance instead of educational performance. Categorization thus may not be a solution to the problem that diversity poses for inter-institutional comparisons and rankings.

Comparisons Based on Resources

What difference do resources make in defining quality? Economists of education draw an important distinction between *tax effort* and *tax capacity*. What they are trying to get at are differences in the resources available to schools, colleges, and universities in jurisdictions that otherwise seem to be comparable. Resources or "inputs" are major factors in most algorithms for ranking colleges and universities. Spending per student, for example, is a measurement of quality based on resources.

Tax capacity is essentially about wealth. But the presence of wealth, however measured, can neither assure nor signify quality. There are two reasons for this. First, a jurisdiction may decide to direct its wealth elsewhere, for example to health care. There are rankings of the "best hospitals" just as there are rankings of the "best universities." The second reason is expressed as tax effort: the extent to which a jurisdiction chooses to tax the wealth available to it.

There might be a tendency to see tax capacity and tax effort as factors that affect only public universities. They affect private universities too. Tax capacity is in the first instance about private wealth, that is, wealth that can be taxed. That is the same wealth that can also be directed to tuition fees, philanthropy, and industrial research. Publicly-funded or guaranteed student financial aid provides what amounts to the working capital of many private colleges and universities.

To the extent that private wealth in the form of tuition fees is part of comparisons and rankings based on resources, rates of private return on private investments – that is, the tuition fee and foregone income – are important. Each year the OECD reports national rates of return on private investments in higher education. Those reports continually show large differences from country to country. The differences are so large – sometimes by factors as high as 2:1 – that they can affect the degree of private wealth directed to higher education.

Taking tax effort and tax capacity into account will not necessarily define quality. In fact, it probably will not. It will, however, advise policy about the prudence and practicality of making investments in order to secure world-class standing. As Philip Altbach (2004) pointed out, being world-class might not be affordable. It also will have a relational value. The *U.S. News & World Report* survey, the *Maclean's* survey, and the Shanghai Jiao Tong University ranking (Liu 2004) all deploy a highly questionable bit of statistical sleight-of-hand. They each assign a value of 100 to the highest-ranking institution in their respective survey. Thus the least imperfect institution becomes the perfect institution. This is done to make the differences among the institutions seem larger and thereby more significant. In terms of tax capacity and especially tax effort, this statistical exaggeration can overstate the putative cost of moving upward on the world-class scale.

Even if the cost of moving upward in a ranking of world-class universities were affordable, two problems might remain. The first highlights the conundrum that difference between reputation and quality causes. Without the statistical inflation of the distance between institutions in ranked order, some institutions may be so close to one another in terms of quality that any investment made to raise an institution's rank would be unwise because the returns would be negligible. The end would not justify the means. A tie would be as good as a win.

The same, however, cannot be said about reputation. Even if statistically unwarranted or even invalid, an upward or downward movement in reputational rank can make a major difference, for example, in the ability to recruit students and faculty. In terms of the screening or signaling function, it can make a difference in the earnings of graduates, which is why alumni continue to be interested in rankings long after the actual quality of their *alma maters* could possibly make a difference to their employment prospects.

The second problem is that rankings can be manipulated in ways that have no relationship at all to investments as costs or as means of improving quality. James Fallows (2001), in a recent analysis of the emphasis placed by elite American universities on early admission, gave the example of the University of Pennsylvania, which was, it appeared, losing its reputation and ability to attract students because of its low placement on various league tables. Through a variety of measures Penn pulled itself up to sixth place in the *U.S. News & World Report* survey. What really happened? One way that a college or university can make itself look more competitive for the purposes of national rankings is to admit more students under early admission. This is a relatively simple matter of arithmetic. For every place in the freshman class that is filled by early admission the ratio of applicant to place is 1:1 and the yield rate is 100 per cent. If, for example, half the class is filled in that way, the number of places to be filled through regular admission, which is what most applicants apply for, goes down by 50 per cent. This tactic drives up the remaining applicant per place ratio and thus increases the overall yield rate as the 100 per cent rate for early admission is blended with the lower rate – sometimes as low as 40 per cent even for elite institutions (Geiger 2000) – for regular admission. Thus the University of Pennsylvania

improved its reputation without making any significant change in its quality or making any major new investments in resources.

Systems *versus* Institutions in Comparisons.

An obvious although frequently overlooked matter of fact in ranking and comparing universities is that institutions are not systems, and *vice versa*. Institutions often have certain characteristics because of the systems of which they are a part. Even institutions that are afforded high degrees of autonomy sometimes are defined in certain respects by the public jurisdictions in which they are located. Absolute autonomy from the state is not feasible (McLendon 2003) and ranking schemes cannot presume that it is.

To a smaller degree, political jurisdictions can affect the performance of private universities, for example, by regulating competition or by controlling access to publicly-funded or guaranteed financial aid. Accreditation is most often found in mixed or predominantly private jurisdictions. Accreditation is, first, much more a national concept than an international concept, the current round of World Trade Organization GATS negotiations notwithstanding. Second, accreditation is about standardization instead of differentiation. Third, it is about minimum standards. It is higher education's version of consumer protection. None of this means that accreditation is useless. But it does mean that accreditation can get in the way of aspirations to be world-class. It also means that, while accreditation can advance the identification of peer institutions within a nation or region, it impedes such identification internationally.

What this suggests is that for any university to be considered even a candidate for world class standing it should have a high degree of autonomy. This not hortatory. It makes a practical difference. This can be exemplified by a contrast of the results of two research projects. In 1997 the Commonwealth Higher Education Management Service surveyed 70 universities world-wide in an effort to measure the degree to which they were affected by state control (Richardson and Fielden 1997). Institutional autonomy and academic freedom were defined and scored separately. In 2003 the Institute of Higher Education at Shanghai Jiao Tong University constructed a ranking of the world's top 500 universities (Liu 2004).

In the Commonwealth Higher Education Management Service study, the region with the lowest state control score for institutional autonomy was North America at 13 per cent. Asia was the highest at 42 per cent. In the Shanghai Jiao Tong study, 63 per cent of the top 500 universities were North American and five per cent were Asian. This does not mean that universities with less independence cannot be of high quality. It, however, does mean that they are not suitable as peers for the purposes of worldwide comparison.

What, for the purposes of selecting peers for comparison, are the hallmarks of autonomy? Burton Clark (1998), Terrence MacTaggart (1998), and Philip Altbach (2004) all present convincing evidence that a high degree of internal self-governance is essential to high performance. When Clark Kerr was chancellor if the University of California at Berkeley he made a point of studying Berkeley's competitors to determine what made them successful. Harvard, Berkeley's major competitor at the time, was notable because " ... as the nation's oldest university, it was the newest in responding the most quickly to out-front developments in the intellectual world" (Kerr 1991, p. 12) Harvard could move quickly because it was self-governing and not burdened by red tape. A decade later Burton Clark observed that quick, nimble internal decision-making was a characteristic shared by successful entrepreneurial universities (Clark 1998).

Self-governance should not be the result of occasional or *ad hoc* policies, no matter how well-intended. Self-governance should be assured by legislation, as should academic freedom. Academic freedom, as a criterion for world-class standing, is not only a political concept. Religious orthodoxy can have the same effects on institutional performance and quality as highly centralized public systems have.

The form of funding – as distinct from the amount of funding – is another telltale sign of the degree of institutional autonomy necessary for selecting peers for world class comparisons. Funding formulas, which are in place in many public systems of higher education, often have powerful standardizing effects. That often is an express purpose of funding formulas (Lang 2005). Nor do they encourage quality and innovation (Ehrenberg 2003). Notably, these are criticisms that are often advanced towards league tables (Kerr 1991; Webster 1992) and accreditation (Skolnik 1986, Ewell 1991).

World-Class or the Curse of Comparison?

To say that the answer to this question is "both" might seem to be an equivocation. It is not. A world class ranking of educational quality is impossible because there are too many permutations and combinations, mandates, and limits imposed by funding to make a reliable selection of peers for comparison possible. To persist in ranking universities internationally in terms of overall quality – that is, attempting to identify the " world's best universities" – is not only an exercise in the production of mis-information, it also is an exercise that will defeat diversity and accessibility. Ultimately, despite the pious claims of *U.S. News & World Report*, *Maclean's*, *Money,* and the *THES,* it will lead to lower quality.

It is, however, possible, if only proximately, to identify slates of peer universities for the purpose of making international comparisons of research quality and productivity (Lang 2000). Comparisons based on research "self defines" an institutional type: the research-intensive university. That definition shortens the list of contending institutions and results in enough conformity to make reliable selections of peers. To understand how this can work we have first to recognize that much of the complexity of developing measures of research quality and productivity can explained by the economic concept of *principal* and *agent.* This is usually called the *agency problem.*

With a few exceptions, principals fund research but do not conduct it themselves. Instead, they purchase research from agents who may be seen as either individual researchers in universities, or as the universities themselves. Therein lies one major complexity. In theory, in terms of the agency relationship, it is the productivity of the agent that should be measured. This suggests that, of the several paradigms that might explain institutional behavioiur in the direction of world class standing, Blau's (1994) is the most plausible: quality is mainly about the performance of faculty.

Funding agencies as principals have different motives and expectations than do researchers as agents. It is difficult, if not impossible, for the sponsors of research to know and, more to the point, evaluate the actual processes of research. But to a considerable degree it is those processes that

productivity is about. Experts who think a lot about agency relationships recommend contractual arrangements that are based on outcomes (instead of process). To the extent that major rankings of universities incorporate performance in research, they measure and rank outcomes. "To the extent" is an important qualifier: the extent often is not great.

Consider this. Each of the two most frequently referred to lists of strategic indicators for higher education in the United States comprises over 100 indicators. On one of the lists (Taylor et al. 1993) research appears only once. On the other (Taylor and Massy, 1996), research appears twice. Both lists contain a "top ten" of indicators. Research appears on neither. Not one of the indicators that involve research could be construed as having anything to do with quality or productivity.

The situation in Canada is not much different. In Ontario the Minster's Task Force on University Accountability (1993) devised 25 indicators of university performance, of which only five involve research. Fourteen of the 25 indicators were identified as "core indicators". Only two of the core indicators involve research. *Maclean's* magazine, in constructing its annual ranking of universities, deploys about 20 indicators of which only three involve research directly, and two more involve research partially. In Alberta the provincial government uses a series of indicators to allocate a portion of its annual operating grants to colleges and universities. Only eight indicators drawn from a possible list of 83 involve research.

In the *U.S. News & World Report* rankings, the weights assigned to research activity – measured in terms of expenditures on research – ranges from zero in the rankings of colleges and business schools to 30 per cent for medical schools. The average is 15 per cent. That the basis of measurement is expense perhaps should not be surprising. As Roger Noll and William Rogerson (1998) observed in *Challenges to Research Universities*, "[A]bout the only aspect of a research program that can be measured objectively is the university's expenditures on inputs." (p. 109) To see the truth in that observation one has only to look at the terms of a typical research grant or research contract. It will be almost exclusively about inputs; little will be said about outputs.

That sounds simple and straightforward until one considers that the process of conducting research is poorly suited to the normal understanding

of the *principal – agent* relationship. Principals, as the funders of research, normally think in terms of outputs. The means by which the outputs are created are essentially unimportant to the principal. To the academic agent as researcher, this is on the one hand welcome and on the other hand a problem. It is welcome because it discourages the principal as funder from interfering with the process of scholarship. Ends count, means do not. This is important to institutional autonomy as an essential criterion for inclusion in rankings of world class standing.

The problem insofar as selecting peers and constructing league rankings is that research is not the only thing that university faculty do. At the University of Toronto, for example, research should, as a rule of thumb, occupy about 40 per cent of a faculty member's time. That is a comparatively high proportion. In North American universities at large the percentage is closer to 30 per cent (Middaugh 1999). So more than one-half of a university professor's time, at least nominally, is not spent conducting research. Indicators of "faculty quality" therefore are not necessarily the same as indicators of "research quality." Again, the quality of research cannot serve as a surrogate for wider quality.

CONCLUSION

If we assume that for some reason one wants to construct, literally, a reliable serial worldwide ranking of universities, the only possible ranking will be a comparative display of research performance only. In this case, the number of universities that can be compared will be relatively small. However, even a ranking based exclusively on research will be problematic for at least three reasons.

First, most existing surveys and rankings place a relatively low weight on research. Second, there are relatively few indices of research performance. Third, different levels and forms of funding for research exist among jurisdictions as well as among disciplines. The heavy reliance in the Carnegie classification scheme on funding from U.S. federal sources explains, for example, why that taxonomy is largely confined to the U.S. There are as well major differences in the amounts of funding available for different research sectors. In some jurisdictions, for example, more than 50 per cent of all public funding for research is earmarked for health science.

Thus the nominal performance of a university with health science faculties will appear to be better than the research performance of a university whose research focus is elsewhere.

Reliable serial ranking of overall educational quality is not possible. The concept of "world-class" compounds the flaws and inadequacies of existing national rankings. If reliable national rankings are not possible, the prospect of a reliable ranking of world-class is even more remote. There are too many different mandates and audiences for universities for a single scale of measurement to produce valid or useful results. Finally, there is no credible or convincing evidence that establishes research performance as an acceptable surrogate for educational performance. So, even if a reliable serial ranking of research performance is constructed, it cannot serve a broader purpose.

While ideally reputation and quality ought to be the same, they are not. Nor are they perceived to be, particularly by students and parents. Reputation, in the form of market branding, is displacing quality as the factor that most influences student choice of university. In that context, both reputation and quality pertain to undergraduate instruction and the services that support it. Thus, just as research performance cannot serve as a surrogate for overall educational performance, educational performance cannot serve as a surrogate for research performance.

Current methodologies – whether commercial or academic – for ranking universities internationally tend inherently toward isomorphism. The American research university is the predominant model. In economic terms, this could be called "look alike" competition. Such competition might be successful in market terms, but in academic terms it works against change and the production of new knowledge. The greatest danger in a ranking of world-class status is not that it would be inaccurate (although it would be), but that it would freeze institutional performance in a monolithic present. This is the curse of comparison. Although there are differences of opinion about the dependence of developing universities on Western models, the isomorphic nature of rankings of world class reinforces that dependence.

Finally, the key to reliable comparisons of universities is the development and use of objective and systematic selections of peer

institutions. Two points must be emphasized to ensure understanding of this conclusion. The first is that a ranking of peers is not about quality or performance. Instead, it is about the degree of difference or similarity among institutions. Once an acceptable degree of similarity has been determined, reliable comparison – particularly for benchmarking – becomes possible. But while this might sound like a conventional ranking, it involves much shorter lists, perhaps no more than ten or fifteen institutions once program, tax wealth, and cost of living are taken fully into account. Program mix is especially important because so many institutional costs and opportunities for research funding are based on the particular array of programs within each university. ❖

References

Acumen Research Group. (1998). *University Applicant Survey.* London, Ontario.

Acumen Research Group. (1999). *University Applicant Survey.* London, Ontario.

Altbach, P. (2004). The costs and benefits of world-class universities. *Academe*, June, 1-5.

Astin, A. (1993). *What matters in college: Four critical years revisited,* San Francisco: Jossey Bass.

Ben-David, J. (1972). *American higher education*. New York: McGraw-Hill.

Berg, I. (1971). *Education and jobs*, New York: Harper Collins, 1971.

Birnbaum, R. (1983). *Maintaining diversity in higher education*. San Francisco: Jossey-Bass.

Blau, P.M. (1994). *The organization of academic work*, Second edition. New Brunswick, New Jersey: Transactions.

Bok, D. (2003). *Universities in the marketplace*, Princeton: Princeton University Press.

Carnegie Council on Policy Studies in Higher Education. (1977). *Selective admissions in higher education.* Washington: Jossey-Bass.

Clark, B.R. (1998). *Creating entrepreneurial universities*. Oxford: Pergamon.

Clarke, M. (2002). Quantifying quality: what can the U.S. news and world report rankings tell us about the quality of higher education? *Education Policy Analysis Archives, 10*(16).

Crane, D. (1972). *Invisible colleges*, Chicago: University of Chicago Press.

Ehrenberg, R. (2003). *Financing higher education institutions in the 21st century*, paper presented at the Annual Meeting of the American Educational Finance Association.

El-Khawas, E. & Massy, W. (1996) Britain's 'performance-based' system, in William Massy, *Resource allocation in higher education*, Ann Arbor: University of Michigan Press, pp. 223-242.

Ewell, P. (1991). The new accountability. *Change*, November-December, pp. 12-17.

Fallows, James. (2001). The early-decision racket. *The Atlantic Monthly*, September, pp. 37-52.

Geiger, R. (2000). Markets and history: Selective admissions and American higher education Since 1950. *History of Higher Education Annual*. Vol. 20, pp. 93-108.

Greene, H.R. (1998). *The select: Realities of life and learning in America's elite colleges*. New York, New York: HarperCollins Publishers.

Hossler, D., Schmit, J. & Vesper, N. (1999). *Going to college: How social, economic, and educational factors influence the decisions students make*. Baltimore, Maryland: The Johns Hopkins University Press.

Just, R. (2002). Rankophile: In defense of the *U.S. News & World Report* college rankings. *The American Prospect*, September 18, 2002.

Kerr, C. (1991). The new race to be Harvard or Berkeley or Stanford," *Change*, May-June, pp. 8-15.

Lang, D. (2000). Similarities and differences: Measuring diversity and selecting peers in higher education," *Higher Education*, Vol. 39, No. 1, pp. 93-129.

Lang, D. (2005). Formulaic approaches to the funding of colleges and universities," in Nina Bascia et al, eds., *International Handbook of Educational Policy,* Dordrecht: Springerr, pp.371-392.

Lang, D. (2005). The political economy of performance funding, in F. Iacobucci and C. Tuohy, eds., *Taking Public Universities Seriously*, Toronto: University of Toronto Press, pp. 226-250.

Lang, D. & Zha, Q. (2004). Comparing universities: A case study between Canada and China. *Higher Education Policy*, Vol 17, No. 4, pp. 339-354.

Lang, K. & Lang, D. (2002). Flags and slots: Special interest groups and selective admission. *Canadian Journal of Higher Education*, Vol. XXXII, No. 2, pp. 103-142

Liu, N.C. et al. (2004). Ranking methodology. Institute of Higher Education, Shanghai JiaoTong University, Shanghai, PRC.

MacTaggart, T. (1998). *Seeking excellence through independence*, San Francisco: Jossey-Bass.

McDonogh, P. (1997). College rankings: Who uses them and with what impact? paper presented at the Annual Meeting of the American Educational Research Association.

McDonogh, P. et al. (1998). College rankings: Democratized college knowledge for whom? *Research in Higher Education*. Vol. 39. No. 5, pp. 513-533.

McLendon, M. (2003). Setting the governmental agenda for state decentralization of higher education. *The Journal of Higher Education*, Vol. 74, No. 5, pp. 479-515.

Marsh, H. & Hattie, J. (2002). The relation between research productivity and teaching effectiveness: Complimentary, antagonistic, or independent constructs?" *The Journal of Higher Education*, Vol. 73, No. 5, pp. 603-641.

Middaugh, M. (1998). How much do faculty really teach. *Planning for Higher Education*, Vol. 27. No. 2, pp. 1-11.

Morse, R. & Flanagan, S. (2002). How we rank schools. in *America's Best Colleges: 2002. U.S. News and World Report*, pp. 9-19.

Noll, R. & Rogerson, W. (1998). The economics of university indirect cost reimbursement in federal research grants," in Roger Noll, ed., *Challenges to Research Universities*, Washington: The Brookings Institution, pp. 105-146.

Ontario Ministry of Education and Training. (1993) *University accountability: A strengthened framework report of the task force on university accountability*.

Pike, G. (2004). Measuring quality: A comparison of *U.S. News* rankings and NSSE benchmarks," *Research in Higher Education*, Vol. 45, No. 2, pp. 193-203.

Richardson, G. & Fielden, J. (1997). *Measuring the grip of the state*, London: The Commonwealth Higher Education Management Service.

Rubenstein, E. (1998). The College payoff illusion. *Outlook*, Fall, pp. 2-7.

Skolnik, M. (1986). How academic program review can foster intellectual conformity and stifle diversity of thought and method. *Journal of Higher Education*, Vol. 60, No. 6, pp. 619-644.

Slaughter, S. & Leslie, L. (1997). *Academic capitalism*, Baltimore: Johns Hopkins University Press.

Smith, A. (1937) *An inquiry into the nature and causes of the wealth of nations*, ed., Max Lerner, New York: Modern Library, Part III, Article II, p. 3.

Spence, M. (2001). Signaling in retrospect and the informational structure of markets. Nobel Prize Lecture, Stockholm, Sweden.

Taylor, B. & Massy, W. (1996). *Strategic indicators for higher education*, Princeton: Peterson's.

Taylor, B., Meyerson, J., & Massy, W. (1993). *Strategic indicators for higher education: Improving performance*, Princeton: Peterson's.

Volkwein, J. F. (1986). Campus autonomy and its relationship to measures of quality. *The Journal of Higher Education*, Vol. 57, No. 5, pp. 510-528.

Volkwein, J.F. (1989). Changes in quality among public institutions. *The Journal of Higher Education*, Vol. 60, No. 2, pp. 136-151.

Webster, G. (1992). Rankings of undergraduate education in *U.S. News & World Report* and *Money*: Are they any good? *Change*, March-April, pp. 19-31.

Winchester, I. (1986). The future of a mediaeval institution: The university of the twenty-first century. in W. Neilson and C. Gaffield, eds., *Universities in Crisis*, Montreal: Institute for Research on Public Policy, pp. 269-289.

Wolf, A. (2002). *Does education matter? Myths about education and economic growth.* London: Penguin.