DOCUMENT RESUME

ED 445 621 HE 033 269

AUTHOR	Lang, Daniel W.
TITLE	Similarities and Differences: A Case Study in Measuring Diversity and Selecting Peers in Higher Education. Professional File, Spring 1999, Number 18.
INSTITUTION	Canadian Society for the Study of Higher Education.
PUB DATE	1999-00-00
NOTE	33p.
AVAILABLE FROM	Canadian Society for the Study of Higher Education, University of Manitoba, 220 Sinnot Building, 70 Dysart Road, Winnepeg, Manitoba R3T 2N2, Canada.
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Access to Education; *Benchmarking; Comparative Analysis; Diversity (Faculty); *Diversity (Institutional); Diversity (Student); Educational Quality; *Excellence in Education; Foreign Countries; Higher Education; Institutional Characteristics; Institutional Environment; Institutional Evaluation; Organizational Climate; *Peer Institutions
IDENTIFIERS	*Canada; *University of Toronto (Canada)

ABSTRACT

Diversity is a policy objective pursued by most higher education systems; at the same time these systems are also concerned about equity of access and the quality of educational opportunity. For a variety of reasons, individual institutions attempt to benchmark themselves against other institutions. Both activities involve measurement, classification, and the selection of peer. Although often addressed apart from each other, diversity and peer selection can be conceptually linked within single scales of similarity and dissimilarity, although existing paradigms that explain diversity may be too simple for reliable peer selection and comparison. A case study of the University of Toronto (Canada) is used to discover the connections between diversity and peer selection, test existing paradigms, and develop a modified methodology that can be used for selecting peers and measuring diversity. Among the study's conclusions are: (1) program cost structures affect institutional cost structures to a large enough extent to be detected in rankings of similarity and dissimilarity and in the measurement of diversity; and (2) of the four principal paradigms--resource dependence, natural selection, competition, and social organization--resource dependence appears to be the most robust in measuring differences in diversity; natural selection and social organization provide better explanations of how diversity develops. (Contains 29 references.) (CH)

The Canadian Society for the Study of Higher Education

Professional File

Spring 1999, Number 18

Le printemps 1999, numéro 18

Similarities and Differences: A Case Study in Measuring Diversity and Selecting Peers in Higher Education

Daniel W. Lang

Department of Theory and Policies Studies, OISE/UT
Division of Economics and Management, Scarborough College
University of Toronto

Abstract

Diversity 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. Individual institutions, for a variety of reasons ranging from accountability to the allocation of scarce resources, attempt to compare or "benchmark" themselves against other institutions. Both activities involve measurement, classification, and the selection of peers. Although customarily addressed apart from one another, diversity and peer selection can be conceptually closely linked within single scales of similarity and dis-similarity. Existing paradigms that explain diversity might be too simple for reliable peer selection and comparison, and might fail to account for all expressions of diversity. A case study is used to discover the connections between diversity and peer selection, test existing paradigms, and develop a modified methodology that can be used for selecting peers and measuring diversity.

Introduction

Diversity

Measuring diversity and selecting peers for comparison are recurrent issues in higher education. Usually they are regarded and discussed as entirely separate topics, each with its own research literature and methodology. Neither, however, is complete or entirely satisfactory. Robert Birnbaum, who has written extensively about diversity in higher education, for example, 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 classification schemes satisfactorily accounts for all institutional characteristics (Birnbaum, 1983; Huisman, 1998).

There are other paradigms. Joseph Ben-David argued that differentiation is the product of competition, and that competition is greatest when colleges and universities are relatively independent (Ben-David, 1972). This would imply a paradigm rooted in organizational behaviour and system structure. From this follows an intriguing paradox: as governments pursue diversity through the construction of more highly regulated and planned systems of higher education they may in practical fact be creating an environment that discourages diversity. This in turn suggests another question: Is it diversity that should be measured or is it the conditions that engender diversity, in this case the level of regulation, which should be measured? Since regulation — which in addition would comprise accountability and the extent to which planning is prescriptive — is an almost exclusively system concept, and since 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.

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

Whatever the paradigm, the scholarship about diversity is aimed principally at two questions: What is diversity and how does it evolve? Diversity is generally accepted as a desirable objective of public policy. From that policy perspective follows another, somewhat more vexing, question which may be asked at both the system level and the institutional level: How does a government know when a sufficient degree of diversity has been realized? How does an individual institution know when it has made a sufficient contribution to diversity? Diversity is neither infinitely valuable, affordable, nor manageable: there can be too much diversity just as there can be too little. This poses problems for at least three critical areas of public policy towards higher education: planning, regulation, and funding. It is at this point that diversity begins to share some characteristics with peer selection.

Peer Selection

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 performance or accountability. 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. Indeed, diversity itself can be benchmarked if a reliable basis of comparability is deployed.

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. As well, and more to the point, accountability based on performance indicators is inherently comparative.

The key to benchmarking and accountability through comparison is not really the indicators or information themselves, but rather the means by which, in regard to benchmarking, an institution, formally through its board of governors, determines its peers for the purposes of comparison. Universities and their boards of governors should be aware of the importance of peer selection and should use it deliberately and formally in various regimes of benchmarking and internal accountability. In regard to accountability and diversity, governments and public agencies should have the same concerns about the basis of comparison, and its potential effect on diversification as well as performance.

Comparisons made ad hoc, either because data are readily available or because comparisons with certain other institutions produce intuitively desirable results, are inherently unreliable and cannot serve accountability and management well. Convenience and politically useful results should not form the basis of peer selection. Neither individual colleges and universities nor systems of higher education can be effectively managed by anecdote. Yet, in the absence of systematic means of determining peers, that is an entirely possible and unfortunately misleading result.

Peer Selection and Diversity: Where do they intersect?

Peer selection is as much an art as a science, and fundamentally involves professional judgement. 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.

An obvious although frequently overlooked matter of fact 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.

Diversity is largely a system concept; it is about groups of institutions defined by political boundaries and about types of institutions defined by various classification schemes. Unless one postulates a virtually infinite number of institutional types, no classification taxonomy can really be about individual institutions, in which case it cannot form a sound and reliable basis for comparing institutions. This ineluctable observation explains why classifications and policies about diversity do not address questions about peer selection, and why peer selection schemes are usually not about diversity.

But if one asks whether or not a given system of higher education is becoming more or less diverse, and whether or not institutions within systems are differentiated, a logical connection to peer selection emerges. Systems can change in two ways: they can add or remove institutions or the existing institutions in them can change. The latter is at least as frequent as the former, and in most Canadian provinces more so. Most classification schemes are not about change, or, more precisely, about degrees of diversity. Peer selection is because it is, in the first instance, about institutions and, in the second instance, attempts to measure institutions more or less continuously.

Think of a continuum with a scale that falls between complete or perfect symmetry among institutions and total dissimilarity or asymmetry. One end of the scale would identify those institutions that for the purposes of benchmarking, performance measurement and accountability can be legitimately and reliably compared with one another. The other end of the scale and the extent to which institutions are distributed along the entire scale would express the degree to which a given jurisdiction or system was diversified. The key point in juxtaposing peer selection and diversity is that in both cases the scale is the same.

Reasons for Interest in Comparative Analysis Using Peer Groups

Strategic Planning

Comparison and emulation are components that are critical in institutional strategic planning. Peer comparisons can provide a basis for the rational evaluation of differences and of similarities among institutions, and of identifying relative strengths, weaknesses, and possible opportunities or niches.

Mission statements are often vague or abstract statements about institutional goals and priorities (Lang & Lopers-Sweetman, 1991). Comparative analysis can help institutions delineate their own identity in more concrete terms. In this regard, such comparisons can be a helpful antidote to external funding and coordination efforts that, deliberately or inadvertently, blur useful distinctions among institutions within a given jurisdiction.

Strategic planning is about a college or university's future aspirations and realistic possibilities. Throughout the research literature on strategic planning there are frequent references to environmental scanning (Bryson, 1988) for the purpose of identifying opportunities, challenges, and the best fits between what the institution is and what its sponsors, users or beneficiaries

wish it to be. Logically, the environment to be scanned for any given institution could have wide and quite indefinite boundaries, so broad and so uncertain as either to defeat scanning or to render it meaningless. By determining its peers, a college or university can give shape to its environmental scanning exercise.

Just as some mission statements are vague and abstract, others are about aspirations, which may or may not be realistic or practicable (Lang & Lopers-Sweetman, 1991). One might think of this means of expressing an institutional strategy as definition by association, whether or not there is a sound basis in fact for the association. So, for example, a university might persistently and publicly compare itself to Harvard to imply that it is somehow like Harvard, and in time and in turn be regarded as being in Harvard's orbit or that it should be funded at that level.

The key, then, to an aspirational approach to determining institutional strategy is to confine or direct aspiration to institutions that, on the basis of comparative data, seem to share a given college or university's mission generally, but appear to be more successful in achieving it.

Alternatively, a given college or university could postulate a different role for itself in the future by defining a "desired institution" containing targets for factors that are potentially controllable by the college or university in the long-term (for example, total enrolment, graduate share of total enrolment, a balance between part-time and full-time balance, library size, instructional program mix) and targets for external circumstances that the college or university might try to have changed (for example, government tuition fee policy), and then use a peer selection methodology to identify those institutions most similar to this "desired institution." The institutions thus identified become a benchmark or milestone against which the college or university can measure its progress.

Although diversity is usually a public policy concern using the idiom of systems of higher education as opposed to that of individual institutions, it can play a role in strategic institutional planning and comparisons that are made in support of it. A quite common strategic planning device is a "strengths and weaknesses" or SWOT inventory which indicates roles for which an institution is most suited (Bryson, 1988). But this device can only be deployed to a certain point in setting strategy and mission. That limiting point is the measure of diversity within the system or jurisdiction within which the given institution is located. If there are a number of other institutions that are already playing the role that the given institution is considering, there may be no niche for that institution to occupy even if it is well suited to the niche. So, institutional plans and strategies sometimes depend on measurements of diversity too.

Evaluation of Institutional Performance

In the absence of absolute standards or frames of reference in higher education for the evaluation of institutional performance, governors and administrators understandably tend to turn to the behaviour of other institutions, either individually or as a group, to establish norms for guidance. Management of higher education is plagued by the "How much is enough?" question. There are no convenient algorithms to determine, for example, what percentage of an institution's budget

should be spent on library acquisitions or how much should be budgeted to produce a given number of instructional hours.

Some "how much is enough" inquiries suggest counter-intuitive results in regard to diversity. For example, if large institutions are more differentiated, and large, complex institutions require greater investments in administration because complexity is more difficult to manage (Blau, 1994), then reducing the cost of administration in the name of efficiency can discourage diversity. So, which performance is more important: administrative efficiency or diversity? This question is more about what should be measured than how it should be measured.

There are a number of quite different ways that administrators and policy-makers attempt to address this question. One of the simplest is to calculate historical averages for various generic categories of expense, and fund all institutions or divisions within an institution on that basis. The averages, once calculated, are then incrementally adjusted for price inflation. Funding for the operation of physical plants is often determined this way. This approach is visibly equitable, predictable and accountable, provided of course that "one size fits all."

Another approach is to presume that in fact one size does not fit all, and that in large complex systems and institutions the extent of experience and knowledge available centrally is not sufficient to make line-by-line decisions about expenditures, a phenomenon that James March calls "limited rationality" (March, 1994). In this case Responsibility Centre Budgeting" is often deployed (Lang, in press). Decisions about allocations under Responsibility Centre Budgeting are deliberately local and program specific, a perspective that inherently discourages comparison, reasoning that local managers know best how to measure performance and allocate resources.

The third approach is comparative benchmarking. A study conducted by the National Association of College and University Business Officers (NACUBO) in conjunction with Coopers and Lybrand was a large-scale benchmarking exercise conducted in the United States and Canada which assembled a very extensive and detailed database that covered virtually every area of institutional activity in higher education. One would have thought that such a study would identify "best practices" among the participating institutions as well as local anomalies that each institution would examine itself (NACUBO, 1993).

But the NACUBO study didn't work that way. Some anomalies were so extreme as to be implausible. Some ostensible best practices, when examined closely, were not portable from one institution to another. There was, in the end, an explanation. Participation in the NACUBO study was voluntary, and it was expensive. A $10,000 fee was charged, as well as the opportunity cost of the staff time needed to assemble the data required from each participating institution. The result was an array of participating institutions that was highly diverse and therefore not conducive to reliable comparison. In other words, there was a peer selection problem.

Prices Paid and Prices Charged

The NACUBO study did demonstrate, however, that large amounts of relevant, definitive data could be assembled across a wide range of institutions. Moreover, the NACUBO study, even on

a preliminary and proximate basis, demonstrated that as far as costs were concerned there were wide ranges of variation, even among institutions that according to Carnegie and AAUP classifications were so similar that they should have had similar cost structures. While, on the one hand, the outcome of the NACUBO study suggests that further comparative studies should be approached with some wariness and skepticism, it, on the other hand, indicates the very considerable potential of such studies if the selection of peers can be undertaken systematically and successfully.

One of the most common applications of peer comparisons — even when conducted casually and anecdotally — is the issue of the prices paid and charged by an institution. Faculty and administrative salaries, tuition and ancillary fees, residence charges, and the cost of purchased goods and services are areas of particular interest.

Fee Ratios

Although some colleges and universities are private and some are public, they all have prices and markets. Marketization is not a phenomenon that is confined to the private sector (Clark, 1998). Moreover, privatization does not necessarily create markets (Marginson, 1997). In many jurisdictions, public policy with respect to tuition fees is changing dramatically. There are many intense debates about tuition fee policy. These debates are often highly political. Comparisons cannot resolve such debates, but they can inform critical decisions about the elasticity of tuition fees as prices.

Both governments and individual institutions should be interested in price elasticity. Governments should be concerned if tuition fees were to have a highly elastic effect on accessibility. They should also be concerned if, by reducing grants while increasing fees, they assume that overall funding will remain approximately the same. If a government were to favour higher tuition fees in order to create and stimulate market behaviour, it should be concerned if fees were inelastic.

Individual colleges and universities not only have to set specific tuition fees, they usually have to set them program by program. Assuming at least some elasticity, setting fees too high would risk unmanageable shortfalls in enrolment. Setting them too low would forego revenue and perhaps imply lower quality programs.

Setting fees by direct comparison is very difficult and unreliable for a number of reasons: fee policy varies significantly from jurisdiction to jurisdiction; there are several educational markets; and only a few institutions actually have international or even national markets. To the extent that fees reflect costs, costs are still variable (as the NACUBO study indicated).

All of this means that the reliable selection of peers is critically important to comparisons of fee levels. It also means that it would be more reliable to compare ratios among tuition fees than to compare fees directly. A ratio in this context would be the percentage by which, for example, the tuition for an MBA program exceeded the tuition fee for a first-year BA. Such ratios could be calculated and compared among both high fee and low fee jurisdictions.

Credibility, Validity, and Control

Credibility, both internal and external, is important. Government funding agencies are often suspicious that *ad hoc* comparisons are contrived to promote institutional self-interest. A systematic, open and detailed process for the selection, and then consistent use, of peers can increase the credibility of comparative results. Internally, peer comparisons can also make possible institutional profiles that provide greater context as opposed to the frequent tendency to assemble isolated bits of polemical comparative data that are sometimes taken out of context.

Although data validity can lead to questions about the appropriateness and reliability of various peer selection approaches, the selection of peers can itself lead to more effective and valid comparisons over time. That is, the development of a stable set or sets of peers enables an institution to focus on a much smaller group of institutions. It can then identify, examine and attempt to rectify differences in definitions and other data comparison problems.

A systematic, pre-determined selection of institutional peers can act as an internal control device. Consideration of comparisons and identification of peers removes the pressure often associated with selecting peers as specific issues arise or as specific decisions are required. Determining peers ahead of time is usually more rational and more credible than selecting them within the political context of a controversial issue. Selecting peers in advance can also add an element of preparedness by assisting an institution in dealing with external requests for data, and in defending against ad hoc peer comparisons developed by other institutions, agencies or the press.

Overcoming Tunnel Vision

Colleges and universities over time may have a tendency to look increasingly inward, either within their own jurisdiction or within themselves. Some degree of complacency or self-delusion with respect to current levels of performance and reputation may result while significant, but unobserved, changes may be occurring in other jurisdictions or at other institutions, some of which might be competitors. Peer selection and comparisons can potentially lead to long-term benefits by shifting an institution's outlook from a relatively internal to a relatively external focus, or at least a focus that engenders greater self-knowledge.

Determining Compensation

Comparisons are part of the warp and woof of collective bargaining throughout the private sector and most of the public sector. Higher education is not an exception. Colleges and universities and the several constituencies within them attempt to make comparisons for several reasons. Employees wish to demonstrate that they are under-compensated in comparison to their putative peers at other institutions. Institutions as employers might wish to demonstrate the opposite. Students refer to comparisons in order to support claims that faculty compensation consumes too large a share of tuition fee revenue. Institutions sometimes deploy comparisons as means of

persuading alumni and funding agencies that additional funds are necessary to maintain salaries at levels that will ensure quality and a competitive position in the academic marketplace.

Because most of these reasons involve at least some degree of self-interest, their credibility depends on objective, consistent, and clearly defined means of selecting peers for comparison. Because in some jurisdictions college and university faculty are employees of a system of institutions or of the state, peer selection that involves compensation must address systems as well as individual institutions.

Peer Selection Methodologies: A Typology

Although not an exact science, there are several methodologies available for determining peer groups among colleges and universities. In the United States, for example, the American Association of University Professors (AAUP), the Carnegie Commission for Higher Education, the National Center for Higher Education Management Systems (NCHEMS), and a few individual states, for example, Washington and Kansas, have developed formal methodologies. Others, like the Maclean's magazine survey in Canada, are less definitive but aim for a similar result. Each uses different criteria but usually includes some subset of the following variables: enrolment, numbers of degrees awarded, programs offered, professional staffing, average salaries, and research expenditures, among others. Some take local geography and demographics into account. A report prepared in 1992 by the Council of Ontario Universities for Maclean's magazine proposed a categorization scheme based on cost structures. So, there are numerous possibilities. Whatever the number of methodologies they can be multiplied by two because the data can be assembled by either institution or program, or both. The differences are potentially significant. For example, certain programs — like Dentistry — may have unique and highly anomalous cost structures that a solely institutional application could mask.

A typology of approaches to developing institutional peer groups is presented in Table 1. The bottom half of the table shows a continuum of options ranging from a judgement-free (statistical approach) to one depending entirely on judgement.

Table 1

Typology of approaches to developing institutional peer groups.

Technique	Cluster Analysis	Hybrid Approach	Threshold Approach	Panel Review
Emphasis	Data plus Statistics	Data plus Statistics plus Judgement	Data plus Judgement	Judgement

It is very important to understand that there can be very large differences between methodologies that organize individual institutions into groups or categories, and then makes comparison among the groups or categories and those methodologies that aim actually to measure the differences or similarities among individual institutions so that they can be compared one to another. With a very few exceptions, the existing methodologies are of the first type: they construct groups of approximately similar institutions according to relatively short lists of characteristics. Once the groups are constructed, the institutions that they comprise are assumed to be identical. These methodologies can assist in comparing jurisdictions in order to measure diversity, but they are unhelpful and even misleading in making other comparisons.

They may not be as reliable as they appear as means of comparing diversity in some circumstances. Many Canadian provinces and several American states have systems of higher education that comprise a lop-sided array of institutional types, for example, by having a single research-intensive "flagship" institution or by having a number of small institutions located mainly to address problems of geographic distribution. Such systems are justifiable, but they are not necessarily comparable as peers despite where their constituent institutions fit in various categorization schemes.

Cluster Analysis

Cluster Analysis is a set of statistical procedures that are designed basically to calculate statistical distance. Alternative ways of making the calculation distinguish alternative clustering methods. Clustering algorithms ensure that the institutions in a given cluster will be more similar to each other, with regard to the variables being evaluated, than the institutions in any other cluster. The approach relies heavily on multivariate statistics and computer processing to manipulate large quantities of institutional descriptors. Other statistical techniques may be used in conjunction with the cluster analysis procedures. Factor Analysis is sometimes used as a step preliminary to Cluster Analysis as a means of incorporating a large amount of data in the peer selection process. Discriminant analysis is used to examine the results of the clustering techniques.

Hybrid Approach

The Hybrid Approach incorporates a strong emphasis on data and input combined with custom designed statistical algorithms for manipulating data. The Hybrid Approach also involves a degree of professional judgement in selection of data and the construction of algorithms. Thus the Hybrid Approach usually involves fewer data than Cluster Analysis because of the pre-selection of data.

Various forms of this approach are conceivable. One such approach is that used by the Kansas Board of Regents to identify peer groups for the six four-year institutions under its jurisdiction (Teeter & Christal, 1987). This methodology was revamped in the fall of 1980 to revise earlier peer selections made by the Kansas Board of Regents, which used these selections as aids in developing funding formulas for institutions in Kansas.

Threshold Approach

The Threshold Approach relies primarily on thresholds and raw data, and depends little, if at all, on statistical methods. It is useful to think of it as a procedure for reducing the universe of institutions until a residue of acceptable ones remains. Although not a pure threshold approach, the National Center for Higher Education Management Systems (NCHEMS) uses a methodology that comes close in practice to such an approach. The Threshold Approach is essentially historical in that it accepts and reinforces data based on fixed performance.

Panel Review

In the Panel Review approach, peer groups are developed primarily through informed judgement, and is based upon the consensus of knowledgeable individuals. Data are used only informally. This approach is commonly used, although descriptions of this approach are difficult to find because of its simplicity and unscientific foundation.

Throughout the former British Common-wealth, "university grants committees" frequently organized institutions into groups or panels for various purposes, including funding.

Reputational surveys are often used either to inform the Panel Review approach or to confirm its results.

De facto or Jurisdictional

A conundrum that confronts several of the paradigms that purport to explain diversification and differentiation in higher education is that the shape and composition of the political jurisdictions in which post-secondary systems function are not themselves the product of, for example, natural selection (Birnbaum,1983) or competition (Ben-David,1972). History, culture, language, and geography are more frequent determinants of political jurisdictions. Any one of these factors can explain certain system characteristics — for example, colleges and universities in remote under-populated areas or, conversely, a congestion of institutions in other areas — that other paradigms cannot.

While other paradigms might be more logical or more theoretically complete, it is neither practical nor reasonable to ignore political jurisdictions in measuring diversity and comparing institutional performance. Thus institutions within a given political jurisdiction and in turn educational jurisdiction are likely to be compared whether or not they would be regarded as similar by any other approach to peer selection.

Some systems are large enough to internalize one of the other approaches, but even then the number of institutions judged to be sufficiently similar for the purposes of comparison might be too small to ensure statistical validity. Other jurisdictions, for example, California, organize institutions into more than one system: universities, four-year colleges, and two-year colleges. And others in the interest of visible equity deploy linear one size fits all funding formulas coupled with local autonomy to promote a modicum of diversity or, at least, an asymmetry between the bases on which funding is allocated and on which it is spent. Whether or not any of these alternatives is commendable, they all exist as approaches that might be taken towards defining institutions that might be considered as peers.

Making a Choice: The Rationale for Using a Hybrid Approach

So just as there are several reasons for wishing to make comparisons among institutions and systems of institutions there are several possible means of making those comparisons. Each offers advantages and disadvantages. Some are more appropriate in certain circumstances than others. One, however, seems to be more commendable than the others.

The Hybrid Approach incorporates the benefits of the Panel Review Approach by requiring the intervention and utilization of expert judgement during the process, as well as at the end, of selecting a final group of peers. The Hybrid Approach has the added advantage of being statistically based, which makes it more objective and thereby more credible than the Panel Review Approach. Consequently, the likelihood of mistakenly selecting an "aspirational" institution as a peer is lower when using the Hybrid Approach than the Panel Review Approach. Such erroneous Panel Review classifications jeopardize the credibility of comparisons, especially in the eyes of third parties like public funding agencies and the press.

Although the Threshold (or NCHEMS) Approach is simpler to use than the Hybrid Approach, the Hybrid Approach has features which make it more attractive despite its relative complexity. It is statistically more sound, and is much more difficult to manipulate, making it more credible to

external agencies and less threatening to potential peers. A major weakness of the Threshold Approach is that it ignores the extent to which institutions miss the value range for a given variable selected by the home institution. The price of this enhanced credibility is a higher degree of logistical complexity. However, only a limited amount of statistical knowledge is needed to comprehend the results of Hybrid Approach.

Cluster Analysis and the statistical techniques that support it, on the other hand, are complex and sophisticated, and require more than a basic understanding of statistics. Although one advantage of the Cluster Analysis approach is that it does not require arbitrary judgements made in advance about the appropriate cut-off points for interval variables as required by the Threshold Approach, considerable judgement is still required to decide both how and where group boundaries will ultimately be drawn, and how to assign weights to the variables entering the analysis.

Cluster Analysis raises other statistical concerns. The manner in which data are standardized can cause problems whereby variables that have the largest variance will have the largest impact on the cluster results, regardless of whether that makes sense substantively. Factor analysis based on samples of fewer than three hundred cases may only have fair reliability.

The technical complexity and abstractness of Cluster Analysis makes it less practical to implement, explain, and understand. Non-statisticians generally have to accept on faith that this approach is appropriate for the selection of peer institutions, and that the human interventions required by these procedures have been reasonable. Cluster Analysis might be more helpful in mapping a universe of institutions, as a government concerned about diversity might wish to do, but, as an approach, it makes less sense when the task is to select a peer group for a particular institution. If Cluster Analysis were used to measure diversity, it would have to be accompanied either by some means of taking national, state, or provincial differences into account or by a weighting scheme to reflect institutional differences that are jurisdictionally determined. In other words, Cluster Analysis would have to be performed twice: once to determine a basis for comparing political jurisdictions, and once to make comparisons among institutions within political jurisdictions previously shown to be similar.

Out of all of the peer selection approaches, the Hybrid Approach is the only one that explicitly takes into consideration the characteristics of the nation, state, province and city in which the candidate institutions are situated. This is desirable because environmental factors are important elements of comparative analyses, for example, ability to pay or cost structures that are based on local costs of living. This recommends the Hybrid Approach to Canadian institutions that wish to select peers among American institutions, and to American institutions in states with relatively few colleges and universities.

The Hybrid Approach makes no preliminary suppositions about institutions by postulating an array of categories and then seeking to determine into which category each college or university should fit. Instead the Hybrid Approach has the potential to reveal and express ranges of similarity.

The Hybrid Approach thus strikes a deliberate and reasonable balance between having

statistical integrity and utilizing professional judgement. It is not so heavily reliant upon judgement that it runs the risk of selecting aspirational institutions as peers or of creating the perception that data have been manipulated to promote institutional self-interest. The major area of subjective judgement — the assignment of selection variable weights — is clearly visible, and thereby open to further review and discussion as necessary. The Hybrid Approach is not so statistically intricate that it is incomprehensible. It is, however, sufficiently elaborate and thorough to discourage the manipulation of results. It permits extensive examination of institutions, particularly with respect to degrees awarded by degree level and instructional program area, and incorporates information on state and provincial characteristics.

A Prototype Methodology

Although there are several theoretical approaches towards the selection of peers, their practical applications have been few in number, and even fewer when applied to measurements of diversity. The methodology and selection of peers described here grew from four similar but separate events, each involving the University of Toronto to some extent.

First was the University's participation in two major data exchanges, the Canadian Universities Data Exchange Consortium (CUDEC) and the American Association of Universities Date Exchange (AAUDE). Comparisons based on peer selection, regardless of theoretical approach, depend heavily on the availability of institutional data. These exchanges provided a wide array of data organized by mutually agreed and recognized definitions

Second was a large-scale benchmarking study sponsored by the National Association of College and University Business Officers (NACUBO). Although NACUBO is an U.S. organization, Canadian institutions were invited to participate in the study, and the Canadian Association of University Business Officers (CAUBO), which is NACUBO's counterpart in Canada, kept an active watching brief on the project. The University of Toronto was a full participant in the project for two years.

Third, in 1991, the Minister of Colleges and Universities in Ontario struck a Task Force on University Accountability chaired by Mr. William Broadhurst, a former president of Price Waterhouse. The task force's final report, which appeared in 1993, made a number of recommendations about performance indicators and how they should be properly deployed. In the task force's judgement, proper use of the indicators depended on definitive mission statements and deliberate and objective identification of peers.

The Broadhurst Task Force, on the one hand, warned against the comparative use of performance and management indicators that were devised in the first instance for purposes of accountability. In particular, the task force expressly explained that none of the indicators that it identified were devised with comparison in mind.

But, on the other hand, the Broadhurst Task Force was neither naive nor unrealistic. It recognized that indicators, once developed and calculated, might be used to make comparisons regardless of

the task force's advice to the contrary. The task force, through a committee that it commissioned to develop indicators, offered two important observations:

> The key to accountability through comparison is not really the indicators. It is the means by which each institution, formally through its board of governors, determines its peers for the purposes of comparison.
>
> Comparisons made willy nilly, either because data are readily available or because comparisons with certain other institutions produce intuitively desirable results, are inherently unreliable and cannot serve accountability well. Convenience and politically useful results should not form the basis of peer selection. *(Task Force on University Accountability, Appendix G,* 1993)

Finally, an Advisory Panel on Future Directions for Post-Secondary Education [Smith Panel] was struck by the provincial government in 1995 and reported in 1997. The panel raised a number of questions about how differentiation among institutions might be measured and promoted, and how distinctive institutional missions and roles might be recognized within a single system of higher education. The panel was also concerned about accountability. Responding to these queries and suggestions required some yardstick by which to express and measure similarities and dis-similarities among institutions.

The University of Toronto therefore had a number of reasons to develop a process for identifying peers and had access to data on which such a process might depend. Those reasons applied both to institutional comparisons and to system comparison based on diversity and differentiation. Some of those reasons, however, posed requirements that went beyond any of the theoretical model methodologies.

After examining the several theoretical peer identification schemes, and favouring the Hybrid Approach, the University of Toronto decided that it should develop that approach further to include four different "slates" of peers: "Base," "Research," "Compensation," "Government Ability to Pay." Each slate would be used in different circumstances but based on the same definitions and data, and organized by program as well as by institution. All data would be drawn from either AAUDE or CUDEC. In addition, data were assembled from various sources on jurisdictional (state or province) characteristics.

That there would be a Base slate could be taken as given. That there should be a Research slate was in part explainable by the role of the University of Toronto, but there were other reasons. Examinations of annual reports of institutional rates of overhead applied to research grants and contracts in the U.S. consistently indicate wide ranges of costs associated with research. Most sources of research funding are national as opposed to state or provincial, in which case the availability of research funding is a factor separate from other factors based on funding.

A Compensation slate was needed for several reasons. Comparisons almost always play a role in labour negotiations about salaries. Salary expense, which is any college or university's single largest cost, can vary significantly among programs. Thus the mix of programs in a given institution can appear to overstate or understate comparative costs unless there is a specific comparison algorithm for compensation. The "compensation" slate is in some respects an

expression of costs of living in different locations. So, for example, all salaries and wages in both the public and private sectors in a large urban area might be relatively high, in which case an unadjusted comparison of higher educational costs would be misleading. A separate "compensation" slate can provide such an adjustment.

Another very frequent use of inter-institutional and inter-jurisdictional comparisons is to lobby government for more funding. Sometimes, perhaps too often, the selection of peers in these comparisons is polemical instead of analytical and objective. Governments know this. The performance of colleges and universities and the degree of diversity in systems of post-secondary education depend heavily on levels of funding. Yet those levels often are not really the result of policies directed specifically at higher education. Instead, they are artifacts of larger policies and circumstances that affect the entire public sector, for example the rise and fall of general revenue. Hence the need for an "ability to pay" slate.

Background: The Logistics of Peer Selection

Canadian Universities Data Exchange Consortium (CUDEC)

In December 1980, the Universities of Guelph, Toronto, Waterloo and Western Ontario and Queen's University took the first steps towards development of a data exchange in response to mutual needs for reliable and consistently defined data about academic units in support of various strategic planning and budgeting. Over the next several years, the scope of the data exchange was expanded to include information on non-academic or non-teaching activities. Institutional participation was expanded to include a number of universities from outside Ontario. In 1986, the Canadian Universities Data Exchange Consortium (CUDEC) was created, and a national steering committee was set up to guide the data exchange process. At its peak CUDEC had fifteen members from seven provinces.

Although data exchange information had been used in the analysis of some divisional resource requests both prior to and since the formation of CUDEC, the University of Toronto's participation in CUDEC was directed mainly to various ad hoc analyses that were usually related in some way to program planning or to the institutional budget processes. There were several reasons for this posture:

i. Individual institutional participation in CUDEC varied from year to year. The result was in some cases databases that were not sufficiently complete for the purposes of time series analysis.

ii. American and European universities are major sources for new PhDs hired into the University of Toronto's tenure stream. Consequently, comparisons to the American labour market for faculty were often more important to salary negotiations than comparisons to other provincial labour markets in Canada.

iii. The University of Toronto, given its breadth, depth, and overall size, had few Canadian peers for the purposes of comparisons that involved certain programs and certain scales of operation.

American Association of Universities Data Exchange (AAUDE)

The American Association of Universities (AAU) is an organization that comprises major research universities in North America. Membership is by invitation. At the time the prototype peer selection methodology was developed, the University of Toronto and McGill University were the only two Canadian members of the AAU.

The AAU Data Exchange (AAUDE) was created in 1973 by interested AAU institution presidents. Its primary purpose was initially to exchange mutually confidential faculty salary and teaching load data, as well as other information of common interest by agreement of institutional representatives, on an annual basis. Since then AAUDE expanded to include a wide range of data and standardized reports.

AAUDE conducts a variety of special studies each year. Participation in those studies often goes beyond the AAUDE membership to include other universities. For example, an academic cost study was undertaken which involved a number of research intensive private universities.

There is also an organization of AAU registrars, called AAUREG. Some comparative data are regularly available through AAUREG. Important examples are data on course and section size.

The raw data supplied to through AAUDE is voluminous. In order to make use of this resource, the University of Toronto decided to generate an annual report that tracked how the university compared, each year and over periods of several years, against AAUDE members with respect to selected institutional statistics obtained through the exchange. These annual reports were forerunners of the sorts of performance indicators subsequently called for by the (Broadhurst) Task Force on University Accountability, and raised in real terms the significance of peer selection.

Task Force on University Accountability

Coincidental to the University of Toronto's review of possible methodologies for selecting peers, interest was mounting on the part of the Government of Ontario over the accountability of Ontario universities for the public funding which they were receiving. In response, a ministerial Task Force on University Accountability was established to undertake a comprehensive review of the accountability practices of Ontario universities and to make recommendations for greater accountability.

In its May, 1993, report to the Minister of Education and Training, *entitled University Accountability: A Strengthened Framework,* the Task Force on University Accountability stated that it considered the governing body of the institution to be the primary and most effective locus of accountability. The Task Force identified two essential accountability functions that should be

the responsibility of the governing body — the approval of policies and procedures covering institutional performance, and the monitoring of them.

To assist it in developing a better understanding of how governing bodies might improve their ability to monitor university activities, the Task Force formally requested that the Committee on Accountability, Performance Indicators and Outcomes Assessment, a sub-committee of the Council of Ontario Universities' Committee on University Planning and Analysis provide detailed advice on benchmarks and indicators that might be used by the individual governing bodies of Ontario universities to improve their ability to hold their institutions accountable. The Committee developed twenty-five management indicators to be employed at the institutional level to inform governing bodies about the activities and performance of the institution.

Although the management indicators were not devised to serve the purpose of institutional comparison or ranking, and the Task Force agreed that they should not be used in those ways, the Committee recognized that governing bodies and other agencies in fulfilling their mandates for accountability might legitimately wish to construct comparative lattices based on these indicators or some sub-sets of them. The Committee pointed out that if any of the management indicators which it devised and which the Task Force recommended were to be used for comparative purposes, it would first be necessary to determine which institutions should be considered as peers for the purposes of comparison.

The Task Force subsequently adopted the Committee's report, included it in its final report, and recommended that universities use the management indicators as part of their obligations for accountability.

For the purposes of objectivity and accountability, and to test the feasibility of the methodology, the prototype methodology was "mapped" to the indicators recommended by the (Broadhurst) Task Force on University Accountability. This was a more significant decision than it might first appear. Most of the classification schemes that are currently in place, as well as methodology proposed by Robert Birnbaum, rely on a relatively small number of variables. Birnbaum, for example, identified six variables: control, size, gender of students, program, degree level, and minority enrolment (Birnbaum, 1983).

The (Broadhurst) Task Force's indicators, however, were wider ranging. This should not be surprising since the task force was concerned with more than diversification and classification. With the exception of minority enrolment; the task force's indicators comprised all of the variables commonly deployed elsewhere, plus a number of others: research grants, research contracts, library resources, international enrolment, faculty awards, student retention and graduation rates, courses offered, instructional workload, balance between full and part-time programs, academic support, and space. Some of these additional variables would have little bearing on diversity, but others would refine the classification, particularly when viewed from the perspectives of Peter Blau or Joseph Ben-David's paradigms.

Adapting the Hybrid Approach to Select Peers

Exchange Rate

Because both the U.S. dollar and the Canadian dollar float a "fundamental equilibrium exchange rate" was set and deployed to align all financial information among institutions. The consistent use of one exchange rate that factored out cyclical variations in currency values was especially important for time series analysis.

Financial Data Adjusted for Geographical Price Differences

Price differences among geographic areas can create significant differences in purchasing power, a condition of major importance in public finance but often overlooked in comparisons and equity considerations. Comparisons of revenues and expenditures lose much of their value if nominal dollar amounts are not adjusted for equal purchasing power. Consequently, the financial data for each AAUDE institution were adjusted using a state Cost of Government Index (COG) developed by the U.S. Department of Education.

The COG reports the market prices and real wages that state and local governments would negotiate for a fixed basket of goods and services purchased for the current operation of their collective public human services, excluding medical services. While not specifically designed for colleges and universities, the COG reflects theoretical minimal prices generally applicable to all public services. For all states, the COG values ranged from a high of 127 for Alaska to a low of 89 for Mississippi. For the 25 states which contained at least one AAUDE member, the COG values ranged from a high of 115 for New York to a low of 90 for North Carolina.

Considerable effort would have to be expended to develop an individual COG value for Ontario, which would be based on the same basket of goods and services as the American COG values. Alternatively, it was possible to use three variables in the peer selection model (population size – 25% weight; urbanization level – 25%; nominal per capita income – 50%) to select the five states that were most similar to Ontario, and then use the average of those states' COG values. Thus, the proxy COG value for Ontario was 98.4 based on Colorado, Florida, Michigan, Ohio, and Washington.

Addition of Library Selection Variables

The University of Toronto placed a high priority on its library system as reflected by a formal budget policy that protected the library acquisitions budgets against budget reductions, price inflation, and currency fluctuation, in other words, ensuring that their real purchasing power was maintained. Given that priority, two selection variables — total library volumes and total library materials expenditures — were added to the peer selection model. This is a good example of the combination of statistical analysis, professional judgement, and selection of data under a Hybrid Approach.

"One-Phase" Selection Process from a Pre-Determined Group

The Hybrid Approach usually follows a "three-phase" selection process. Taking the State of Kansas as an example, the first phase involved the identification of the 33 states that were most similar to Kansas in terms of population, urbanization level, nominal per capita income, and high school attendance patterns in higher education. The second phase reduced and grouped the number of institutions within the remaining 33 states using institutional characteristics such as ownership (public versus private), institutional type, number of doctoral programs offered, and the size of the city within which the institution is located. The third, and final phase, then determined the similarity of the remaining institutions to the home institution with respect to enrolment, funding and expenditure patterns, and degrees awarded.

The proposed peer selection methodology for the University of Toronto used a "one-phase" selection process given the recommendation to select its peers from a predetermined candidate group, the major research universities that were members of the AAUDE. Three of the six state characteristic variables used in the first phase of the Hybrid Approach, for which Ontario information exists, were considered simultaneously in the proposed Toronto methodology with the enrolment, funding and expenditure pattern, and degrees awarded information. That is, the six state characteristic variables in the Hybrid Approach were used only as an initial screening device and did not contribute towards the total similarity score for each institution whereas the three characteristic variables for Ontario and the states in the proposed Toronto methodology were not used as a screening device. Instead they contributed a portion of the overall similarity score for each institution.

Because the membership of the AAU is essentially a combination of self-selection and invitation, the University of Toronto also undertook a separate state similarity analysis using information on all 51 states. Only five out of the 38 AAUDE members are not situated within the 33 states calculated as being most similar to Ontario, four from California and one from New Jersey. California is very dissimilar from Ontario, and all other states, due to its large total population of 29.8 million while New Jersey is dissimilar from Ontario, and almost all other states, due to its high per capita income. These five institutions were excluded from the peer selection analysis, however, given that the state/provincial characteristic variables, although appropriate factors for the determination of peer institutions in a broad sense, were relatively not the most important selection variables overall.

Although sharing similar research missions, AAU institutions still varied according to such characteristics as institutional size, enrolment, financial resources, library size, state or provincial characteristics, and program mix as reflected by degrees awarded.

Four Proposed Slates of University Peers

In some jurisdictions, governing agencies use peer selection models to select one group of peer institutions for each institution within the jurisdiction. Even within a given institution, however, a case can be made for different slates of peers depending on the particular comparisons that a board of governors might wish to make for the purposes of accountability. A variety of slates was possible. The University of Toronto deployed four slates, which are outlined by Table 2.

The four slates were differentiated by the relative weights assigned to the peer selection or data input variables as follows:

- The selection variables were conceptually grouped into three categories: State/Provincial Characteristics, Enrolment/Financial/Library, and Degrees Awarded. The total residual weight between the latter two categories was split 50:50 once the weight for the first category has been determined.
- For the Base and Compensation slates, the total weight assigned to the degrees awarded category was then equally distributed among the selection variables for each of the four degree levels. That is, the degrees awarded category was assigned a high weight in total, but a neutral position was taken with respect to the relative importance of each degree level to the selection of a peer group. The Research slate assigned higher weights to the master's and doctoral degrees awarded selection variables. The degree level weights for the Government Ability to Pay slate reflected the actual distribution of degrees conferred in 1987-88 by degree level expressed in government funding units.
- For the Research slate, higher weights were also assigned to the research expenditures, graduate and first professional share of full-time equivalent enrolment, and library selection variables.
- For the Compensation slate, higher weights were assigned to the urbanization level, per capita income, graduate and first professional share of full-time equivalent enrolment, tuition and fees revenue, and restricted funds revenue.
- For the Government Ability to Pay slate, higher weights were assigned to the state or provincial characteristics, and tuition and fees revenue selection variables.

Table 2
Peer Selection Variable Weights

Selection Variable	Base State	Research State	Compensation State	Gov't. Ability to Pay State
ENROLMENT:				
FTE Enrolment	5.0	2.0	0.0	5.0
Head Count Enrolment	3.0	0.0	0.0	0.0
Graduate & First Professional as a % of FTE Enrolment	8.0	12.0	16.0	5.0
	16.0	14.0	16.0	10.0
FINANCIAL:				
Current Fund Expenditures	4.0	2.0	2.0	0.0
Tuition and Fees Revenue as % Current Fund Revenues	4.0	2.0	8.0	15.5
Instruction Expenditures	4.0	2.0	2.0	0.0
Restricted Funds Revenues	4.0	2.0	8.0	2.0
Research Expenditures	4.0	6.0	2.0	0.0
Research Expenditures as % Current Fund Expenditures	4.0	8.0	2.0	0.0
	24.0	22.0	24.0	17.5
LIBRARY:				
Library Volumes	4.0	8.0	0.0	0.0
Total Library Materials Expenditures	4.0	6.0	0.0	0.0
	8.0	14.0	0.0	0.0
STATE/PROVINCIAL:				
Population Size	0.0	0.0	0.0	15.0
Urban as % Population	2.0	0.0	10.0	5.0
Per Capita Personal Income	2.0	0.0	10.0	25.0
	4.0	0.0	20.0	45.0
DEGREES AWARDED:				
Bachelor	12.0	7.5	10.0	13.9
Master's	12.0	15.0	10.0	7.8
Doctoral	12.0	20.0	10.0	2.5
First Professional	12.0	7.5	10.0	3.2
	48.0	50.0	40.0	27.5
TOTAL	100.0	100.0	100.0	100.0

The process of peer selection under the adapted Hybrid Approach

The actual work of identifying peer institutions and assembling slates of institutions was very time-consuming. At any given time, as many as three professional staff from the Office of the Vice-Provost and Assistant Vice-President, Planning and Budget at the University of Toronto were working full-time on the project. They were Ken DeBaeremaeker, Nasreen Jivraj, and Anthony DiFelice. The successful outcome of the project depended in large part on their intelligence and ingenuity.

The first step was the incorporation of a wider range of variables. Conceptually this was not a difficulty, but it did complicate the logistics of data definition and collection.

Next, all of the other institutional members of AAUDE were examined. Four institutions — Brandeis, California at Irvine, California at San Francisco, Columbia — were eliminated because complete information was unavailable for each. The remaining thirty-seven institutions, which were referred to as the "candidate group," were screened by similarity to the University of Toronto with respect to enrolment, funding and expenditure patterns, library volumes and materials expenditures, state or provincial characteristics, and degrees awarded.

A mean and a standard deviation were calculated for each selection variable from which a z-score[1] was generated for each institution. Each candidate's z-scores are compared to those of the University of Toronto by taking the absolute value of their differences. The results of this process are referred to as "comparison scores."

To compare degrees conferred, a matrix of degrees awarded by instructional program area and by degree level (bachelor, master, doctoral, and first professional) was generated for each institution. From this pool of matrices, a mean and standard deviation was derived for each cell of the matrix, from which a z-score and comparison score were calculated for each cell of each institution's matrix. Each institution's instructional program area comparison scores were then aggregated by degree level and divided by the number of instructional program areas where degrees were awarded by both the candidate peer institution and the University of Toronto plus the number of instructional program areas in which degrees were not awarded by either the candidate institution and the University of Toronto. This resulted in four comparison scores per institution, one for each degree level.

The reason for discriminating among programs that were offered by both institutions, only one, or by neither was the knowledge gained from previous NACUBO and CUDEC analyses that had indicated that some programs — for example, Dentistry — had highly anomalous cost structures that could have a powerful effect on comparisons. While that effect might be statistically noticeable in institution-to-institution comparisons, they might be masked when systems were compared to one another.

All comparison scores (c) were then standardized using the formula $X = 10 + 5c$. Since z-scores commonly range between -3 and 3, this conversion caused the comparison scores to become non-

[1] z-score = (raw datum – mean for variable) / standard deviation for variable

negative with broader ranges. In the case of degrees awarded, however, only standardized comparison scores were provided for each institution, one score for each degree level.

The cells of the matrices for the five institutions not awarding any degrees at the first professional level — Carnegie-Mellon, Maryland at College Park, MIT, Michigan State, Pennsylvania State — were excluded from the above computations for the first professional degrees awarded selection variable because they would necessarily have had undefined input values. The standardized comparison scores for those institutions' first professional program variables were artificially set at 10.5, or just above the highest standardized comparison score among all the institutions that award first professional degrees. That is, those institutions awarding no professional degrees were at most no more similar than the least similar institution that awarded professional degrees.

Weights (totaling 100) were applied to the standardized comparison scores of the selection variables. The scores thus weighted were summed to create similarity scores. The institutions were then rank-ordered by similarity score. These rankings then served as a valuable aid in selecting a final set of peer institutions.

It should be noted that the above methodology always results in a similarity score of 1,000 for the University of Toronto because all of its comparison scores, by definition, must equal zero. At the same time it is important to understand that a low score is just as instructive as a high score because under the prototype methodology the fundamental objective is to measure ranges of institutional similarity. The wider the range, the greater the diversity. The higher the comparison score, the closer the similarity among potential peers. Depending on the distribution of scores, the methodology could suggest de facto systems within jurisdictions that do not formally or intentionally seek to differentiate among institutions (as was the case of Ontario and the University of Toronto).

Calculation of Comparison Scores for Each Degree Level

A " comparison score" was calculated for each degree level by dividing the sum of the comparison scores for each instructional program area by a count or CNT value equal to the number of program areas for which degrees were awarded by both the candidate institution and the University of Toronto.

One effect of the above calculation was to magnify to varying degrees: similarity based on comparable program offerings, similarity based on lack of program offerings, and dissimilarity based on different program offerings. In isolation, such an effect might have been desirable. The level of magnification was significantly high, however, even for institutions with many comparable program offerings based on the fact that a majority of the 50 instructional program areas are not offered by the AAUDE institutions, even at the bachelor degree level. For example, the University of Toronto awarded degrees in only 21 instructional program areas at the bachelor degree level, 22 program areas at the master's level, 19 program areas at the doctoral level, and 2 program areas at the first professional level. These numbers represented the maximum CNT

values, given that CNT was equal to the number of instructional program areas where degrees were awarded by both the candidate institution and the University of Toronto.

The principle, which was first adopted by the University of Kansas, of excluding instructional program areas from the CNT value where degrees were not awarded at the degree level in question by both the candidate institution and the University of Kansas seemed questionable. Although mission statements are rarely expressed in such a fashion, institutions may be as similar in terms of what they do (programs offered) as in terms of what they do not do (programs not offered, sometimes because of government regulation).

A detailed review of the comparison score and similarity score calculations revealed that a combination of instructional program areas from the CNT value where degrees were not awarded by both the candidate institution and the University of Toronto, the formula used by the Kansas Board of Regents to standardize the comparison scores, and the proposed weights for the degrees awarded selection variables had a strong arithmetic effect resulting in total similarity scores that created an impression that certain institutions were less similar to the University of Toronto than they in fact were. The CNT value therefore was changed to equal the number of instructional program areas where degrees were awarded at the respective degree level by both the candidate institution and the University of Toronto plus the number of instructional program areas where degrees were not awarded by both the candidate institution and the University.

Standardizing Comparison Scores

The Kansas Board of Regents standardized the comparison score (c) for each selection variable using a formula $X = 50 + 10c$. That is, for presentation purposes the comparison scores were magnified by the formula over a broader range. (Although such standardization formula would not change any institution's relative position vis-à-vis the home institution for each of the selection variables under examination, the necessity of using the standardization formula.) In particular any coefficient values of as large as 10 were questionable because they could result in total similarity scores that left an impression that institutions were less similar to one another than they in fact were. It was decided therefore to keep the standardization formula, but change it to $X = 10 + 5c$.

Classification of Degrees Awarded by Instructional Program Area

The AAUDE institutions report their degrees awarded information using the Classification of Instructional Programs (CIP) developed by the U.S. Department of Education's National Center for Education Statistics (NCES). The CIP is used in all NCES surveys and is the accepted U.S. Government standard on programs for education information surveys.

The University of Toronto's degrees awarded information was mapped to fit the CIP scheme. The enclosed glossary contains the definition of each degree level: bachelor, master, doctoral, and first professional. An important note: this was not difficult to do, nor was there any indication that it would have been difficult for other non-AAU institutions to do.

The Results: Four Slates of Peer Institutions

While there was a conceptual basis for identifying and seeking to calculate four separate slates of peer institutions, it could not be taken as given that the comparison scores, when calculated, would actually indicate statistically significant differences among institutions by slate. In other words, each slate might have comprised the same institutions in the same ranked order. That in turn could have meant that diversity among institutions and among post-secondary systems was a problematic concept to express by classification.

The results, however, were as anticipated; there were indeed differences among the slates, as Table 3 indicates.

Table 3
University of Toronto Peer Institution "Top-Ten" Results ranked in ascending order to similarity to the University of Toronto

BASE SLATE	RESEARCH SLATE
Arizona	Arizona
California, Berkeley	California, Berkeley
Illinois, Urbana-Champaign	California, Los Angeles
Michigan	Michigan
Minnesota	Minnesota
North Carolina, Chapel Hill	North Carolina, Chapel Hill
Ohio State	Ohio State
Rutgers	Rutgers
Texas, Austin	Texas, Austin
Washington	Washington
COMPENSATION SLATE	**ABILITY TO PAY SLATE**
Arizona	Arizona
California, San Diego	Illinois, Urbana-Champaign
Florida	Michigan
Illinois, Urbana-Champaign	Michigan State
Kansas	Minnesota
Missouri	Missouri
Ohio State	North Carolina, Chapel Hill
SUNY, Buffalo	Ohio State
Texas, Austin	Texas, Austin
Washington	Washington

At this point it is critically important to recognize the crucial role that the selection variable weights played in the analysis. Changes in the weighting resulted in changes in the similarity scores. The weights were the connection between the statistical dimension of the Hybrid Approach and its judgmental dimension. While this characteristic of the Hybrid Approach is not

difficult to understand in theory, it is difficult to deploy in practice. The weights were in effect a missing link that solved this problem.

Although there were changes in the ordinal rankings, the overall "top-ten" results for the Base and Research slates differed by only one institution each. In a sense, there were two research slates, each with a different emphasis on research intensity. That is, the Base slate by itself is in some ways a research slate given that it was selected from a pre-determined group of primarily public, primarily research universities. The Research slate was created by assigning higher weights to the graduate and first professional enrolment share, research expenditure, library, master's degrees awarded, and doctoral degrees awarded selection variables.

While the Base and Research slates were very similar in terms of composition, they were less similar in terms of ranked order. This suggests that for the purposes of constructing groups of institutions for comparisons of diversity among systems the array of slates could be different from the array that an individual college or university might wish to deploy for the purposes of peer selection. The methodology, however, would otherwise be the same in both cases.

Four institutions — Arizona, Ohio State, Texas at Austin, Washington — were within the "top-ten" peer group for all four of the proposed slates of university peers. Four other institutions — Illinois at Urbana-Champaign, Michigan, Minnesota, and North Carolina at Chapel Hill — were "top-ten" peers for three of the four proposed slates.

That there was a fixed number — ten — in each group was arbitrary for validation and demonstration purposes. Final peer groups could have included a larger (or smaller) number of institutions given that the differences in similarity scores between the tenth and immediately following institutions were not statistically great. In all cases, the raw data from which the similarity scores were generated were reviewed before final judgements were made about each of the proposed slates peers in order to determine whether the cut-off point should be moved lower or higher for each list of institutions sorted by similarity scores.

The ranges of comparison scores varied among the four slates from 1,347 to 1,287 in the "top ten" category, and from 1,847 to 1,669 overall. A score of 1,000 represented a perfect match with the University of Toronto. No private institutions ranked high in terms of similarity. That outcome was not surprising given the significance of scale in the methodology (in fact, in all the methodologies). The University of Toronto is a very large, multi-campus institution. Among AAU members, private universities all were among the smaller institutions. That also explains why other Canadian universities would not rank high in terms of similarity.

Conclusions: What does the case study tell us about peer selection and measuring diversity?

Program cost structures can effect institutional cost structures to a large enough extent to be detected in rankings of similarity and dissimilarity and in turn in measurements of diversity.

"Program" is among the most problematic terms in the higher education lexicon, especially within the context of diversity. Sometimes the concept of program is expressed as levels of credential conferred: undergraduate, master's, doctorate (Birnbaum, 1983; Rawson, Hoyt, & Teeter, 1983; and Teeter & Christal, 1987). In other cases "program" means disciplines and fields of study, so for example physics is a program regardless of degree level (Huisman, 1998). And in other cases the mode of delivery is regarded as a "program" characteristic (Jones, 1996).

Any one or all of these understandings of what "program" connotes might reasonably be taken into account in measuring and expressing diversity. Most approaches use the first: "program" means degree offered. However, in constructing the peer selection methodology in the University of Toronto case study it became evident, particularly from the research slate, that the definition of program, which made the most difference in terms of resources, was organizational. A faculty, school, or department was a "program."

On reflection, the organizational concept of program makes sense. Expenditures within postsecondary institutions are usually assigned to programs as organizations, that is, to faculties or departments. In some cases, revenue too is attributed to programs as organizations (Lang, in press). Real program budgeting (PPBS) has been tried in higher education but with little success (Massy & Hopkins, 1996).

Moreover, the single largest area of expense in higher education is salaries. That was a principal reason for the University of Toronto's decision to construct a separate compensation slate. When comparisons are based on compensation, two additional comparative factors come into play: the distribution of faculty by rank (Terenzini, Hartmark, Lorang, & Shirley, 1980) and the mix of programs (Simpson & Sperber, 1988). Both of these factors use the organizational idiom for program.

What this means for the selection of peers and the measurement of diversity is that the organizational definition of "program" is at least as important as the more commonly used degree offered definition, and that, even when the objective is to compare a diversity among systems of higher education, taxonomies and other classification schemes should begin at the program level and build up from there.

Of the four principal paradigms — resource dependence, natural selection, competition, social organization — it would appear that resource dependence is the more robust in measuring differences in diversity, whereas natural selection and social organization might provide better explanations of how diversity develops.

Although other applications of the Hybrid Approach have taken jurisdictional characteristics into account and weighted them (Rawson, Hoyt, & Teeter, 1983), none has sought to determine ability to pay except in terms of per capita income. But there is little evidence that per capita personal income determines public spending on higher education. There are some jurisdictions in which funding for colleges and universities is determined as a fixed share of either government revenue or government expenditure (Ziderman & Albrecht, 1995). There are, however, numerous

factors that come between per capita personal income and total government revenue and spending.

Among the more obvious intervening factors are funding formulas, subsidies to students, research and development policy and spending, rates of matriculation from secondary school, and other priorities for public spending. Even revenue from tuition fees, which would appear to be directly related to per capita personal income, is significantly determined the distribution of personal incomes and the availability of subsidies to students (Lee, 1987).

The construction of the ability to pay slate indicated, first, that ability to pay is a powerful and independent factor in measuring institutional similarity and dis-similarity. Second, it indicated that the measurement of ability to pay depended more on the amount of general revenue available to a government for allocation, and on the policies and means by which general revenue is allocated, than on gross personal wealth.

There are significant differences among institutions which other commonly used categorization schemes fail to detect.

Consider the implications of the following observation made possible by the case study and in particular the use of separate slates of institutions for comparison: under either Carnegie Commission or AAUP classification scheme — the two most commonly used taxonomies — all of the institutions in the case study would have fallen into a single category, yet the case study statistically validated at least four different slates of institutions. One implication is that, because all of the institutions would have been located in a single category, they would be assumed to be identical for the purposes of comparison and of measuring diversity. But the variations among the slates indicate that differences among institutions — for example, in salaries or in research intensity — do not "average out" and become statistically negligible.

Diversity is more than descriptive. The fact that four slates could be statistically validated suggests that for each policy objective for diversity there should be a separate comparison and formation of peer groups.

Because there are real differences among otherwise putatively identical institutions which are more than statistical wrinkles that can be ironed out, systems of higher education, like individual institutions, should be more concerned about peer selection. While institutional size, degrees offered, and program mix will perhaps continue as the predominant expressions of diversity among systems of higher education, other expressions can have useful roles to play. For example, to the extent that resources determine quality, regardless of the types of institutions involved, ability to pay and compensation (which in turn involves the mix of disciplines and the mix of ranks) become vital factors for comparison. For another example, the organization and cost of research varies so considerably from disciplinary area to disciplinary area that diversity in research and advanced graduate study (as measured by the doctoral and doctoral stream master's programs and enrolment) cannot be adequately represented by existing taxonomies. If that

proposition were not true the research slate in the case study would have been the same as the other slates.

The same methodology can support measurement of diversity as well as the selection of peers.

The range of variation among comparison scores overall was quite similar for three slates — Base, Research, and Compensation — and quite different for the fourth, Government Ability to Pay. In the "top ten" The range of variation between the Base and Research slates was minor but the Compensation and Government Ability to Pay slates were quite different from the Base and Research Slates, and from one another. Within each slate the range of variation was significant.

These results indicate two things. First, individual institutions need to take care in selecting peers. Intuitive, *ad hoc,* and aspirational selections are not reliable. Second, the commonly deployed categorization taxonomies mask differences that could be significant in comparisons of diversity among jurisdictions. For example, the Government Ability to Pay slate is the most different among the four slates. While all jurisdictions would wish to increase their public and private wealth, few would have much ability to control or force such an outcome. Thus differences in Government Ability to Pay are as unavoidable as they are significant. But neither observation would be fully apparent from the existing classification schemes.

The comparisons score and in turn the ranked order slates are obviously applicable to the selection of individual peer institutions. The peer selection methodology could also apply to systems. Diversity could be represented by a desired range of comparison scores instead of by aggregations of institutional types. Like the University of Toronto in the case study, jurisdictions might wish to deploy the methodology with slates, and perhaps add new slates. For example, accessibility is largely a system concept. A slate that weighted more heavily the variety and capacity of degree programs that could be entered directly from secondary school might be of particular interest to some jurisdictions.

References

Council of Ontario Universities. (1993*). Report of the Committee on Accountability, Performance Indicators, and Outcomes Assessment to the Minister's Task Force on University Accountability.* March 12, 1993. Council of Ontario Universities.

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

Ben-David, J. (1977). *Centers of learning.* New York, NY: McGraw-Hill.

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

Blau, P.M. (1994). *The organization of academic work,* 2nd edition. New Brunswick, NJ: Transactions.

Brinkman, P.T. (1987). Effective institutional comparisons. *New Directions for Institutional Research, Vol. 53.*

Brinkman, P.T., & Teeter, D.J. (1987). Methods for selecting comparison groups. *New Directions for Institutional Research, Vol. 53.*

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

Cleaver, G.S. (1981). *Analysis to determine a ranking in similarity for institutions in higher education.* Lawrence, KS: University of Kansas, Office of Institutional Research and Planning.

Huisman, J. (1998). Differentiation and diversity in higher education. In J.C.Smart, ed., *Higher education: Handbook of theory and research, Vol. XIII.* New York, NY: Agathon.

Huisman, J., & Morphew, C. (1998). Centralization and diversity: Evaluating the effects of government policies in the U.S.A. and Dutch higher education. *Higher Education Policy, 11*(1).

Jones, G.A. (1996). Diversity within a decentralized higher education system: The case of Canada. In V.L. Meek et al., eds., *The mockers and the mocked: Comparative perspectives on the differentiation, convergence and diversity in higher education.* Oxford, UK: Pergamon/IAU.

Lang, D. (in press). Responsibility centre budgeting and responsibility centre management in theory and practice. *Higher Education Management.*

Lang, D., & Lopers-Sweetman, R. (1991). The role of statements of institutional purpose. *Research in Higher Education, 32*(6).

Lee, J. (1987). *The equity of higher education subsidies.* College Park, MD: University of Maryland National Center for Postsecondary Governance and Finance.

March, J. (1994). *A primer on decision making.* New York, NY: The Free Press.

Marginson, S. (1997). *Markets in education.* St. Leonards, NSW: Allen & Unwin.

Massy, W., & Hopkins, D. (1996). Lessons from health care. In W. Massy, ed., *Resource allocation in higher education.* Ann Arbor, MI: University of Michigan.

Middaugh, M. (1998). How much do faculty really teach. *Planning for Higher Education, 27*(2).

Mims, S., & LeLong, D. (1987). The Michigan Peer Institutions Information Study. *New Directions for Institutional Research, 53.*

NACUBO with Coopers & Lybrand,and Barbara Shafer & Associates.(July, 1993). *Benchmark 92, Final Report.*

Ontario Ministry of Education and Training. (1993). *University accountability: A strengthened framework.* Report of the Task Force on University Accountability, May 7.

Rawson, T.R., Hoyt, D.P., & Teeter, D.J. (1983). Identifying 'comparable' institutions. *Research in Higher Education, 18*(3).

Research Associates of Washington. (1989*). State Profiles: Financing public higher education 1978 to 1989,* Twelfth Edition.

Simpson, W., & Sperber, W. (1988). Salary comparisons: New methods for correcting old fallacies. *Research in Higher Education, 28*(1).

Terenzini, P., Hartmark, L., Lorang, W. & Shirley, R. (1980). A conceptual and methodological approach to the identification of peer institutions. *Research in Higher Education, 12*(4).

Teeter, D.J., & Christal, M.E. (1987). Establishing peer groups: A comparison of methodologies. *Planning for Higher Education, 15*(2).

Whiteley, M.A.,& Stage, F.K. (1987). Institutional uses of comparative data. *New Directions for Institutional Research, 53.*

Ziderman, A., & Albrecht, D. (1995). *Financing universities in developing countries.* Washington, DC: Falmer.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

☑ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").