

## Rankings of Canadian Universities: Pitfalls in Interpretation

STEWART PAGE\*

---

### Abstract

*A critical perspective is presented in regard to rankings of Canadian universities by Maclean's magazine, November 11, 1993. Some brief comparisons are also made in regard to the 1992 rankings and data. Several pitfalls in the ranking procedures, and the results of some correlational analyses of the ranking data, are outlined. A brief summary of comments and some implications are presented, bearing on the wider issue of 'public' university accountability and also on the practical issue of students' choice of universities.*

---

### Résumé

*Point de vue critique offert dans la revue Maclean's sur le palmarès des universités canadiennes, publié le 11 novembre 1993. Brèves comparaisons à partir du rang attribué et des données recueillies en 1992. Survol des pierres d'achoppement du système d'attribution du rang ainsi que des résultats découlant de certaines analyses corrélationnelles. Résumé concis des universités dites «publiques» et, d'ordre plus pratique, le choix d'universités à la disposition des étudiants.*

---

\* University of Windsor. The author thanks Dr. Arthur May, President of Memorial University, Dr. Robin Farquhar, President of Carleton University, and Dr. Shelagh Towson, Department of Psychology, University of Windsor, for comments rendered on a previous draft of this article. Opinions and conclusions expressed herein are, however, solely those of the author. Thanks are also extended to Rosie Page for editorial assistance, and to the editor and two anonymous reviewers of the *Canadian Journal of Higher Education* for comments which were helpful in undertaking revisions.

In its November 11, 1993 issue (pp. 29-73), *Maclean's* magazine (MM) published its third annual rankings of Canadian universities. The expressed intentions of this exercise were to inform the public, to help a university to "clarify its own vision," and ostensibly to give students a "critical tool," indeed the "definitive road map," with which to judge universities' strengths and weaknesses. In asking "What, besides a piece of paper, does a university degree really provide?" the magazine considers higher education in the context of its instrumental value, that is, in terms of its potential to provide access to "today's job market." To render its analysis of higher education parameters more intelligible, MM summarizes it with pop metaphors. Thus, in contrast to others, Mt. Allison attracts scholars who care, Simon Fraser is the one with open doors and open minds, which have propelled it to the top, where it is now perched at the summit. Also, McGill is said to have the right stuff.

This paper outlines some pitfalls in the ranking and overall statistical approach taken by MM. Second, some practical implications of these pitfalls, as they bear on the issue of student choice of university are noted briefly. It is hoped that these perspectives on the MM approach might serve to attract and to help focus future debate on the wider issue of how university evaluation and accountability should best be addressed in a public forum such as that provided by a mass circulation magazine. It should be noted, at the outset, that the present discussion is based on the 1993 rankings and procedures and contains no explicit reference to the 1994 data published by MM in the fall of 1994. Regarding its procedures for 1994, MM published a somewhat more complete account of the ranking data and component parts thereof and more exact descriptions of its procedures compared to those for 1993. In the latest exercise, however, MM appears to have retained its basic philosophy of simple summation and conversion of point totals to ranks, with these again leading to creation of a linear (vertical) "ranking" of universities, based on selected "criteria."

MM is equivocal about how well the universities cooperated in supplying the evaluative data. It says they should be complimented for being a "brave example in public accountability," and "preserving a tradition of excellence," yet it also says that obtaining data constituted a "battle for the facts," that the universities have responded "at a snail's pace," and that they possess a "deep unease over accountability." It might be noted that universities typically employ their own student-driven evaluation systems with results made available to students. Courses and their professors (who represent much of how a university is experienced by students) are therefore generally accountable on a regular basis. The rankings approach presently does not include or incorporate these data or indices.

## Measures Used

MM classified universities into Medical/Doctoral ( $N = 15$ ), Comprehensive ( $N = 13$ ), and Primarily Undergraduate ( $N = 23$ ) categories depending on its judgements concerning the extent of a university's involvement with graduate programs and research.

Data based on a 14-page questionnaire sent by MM to the universities in July 1993, were compiled according to the following six measures: Student Body (comprised of six indices of student ability, such as the grade average of incoming students); Classes (four indices of class size and "quality"); Faculty (four indices of faculty calibre, rank, and grant record); Finances (three indices of budget, student services, and awards); Library (three indices of collections); and Reputation (two indices, based on alumni support and on a reputational survey sent to senior university officials and chief executive officers of Canadian corporations). Based on a preliminary point allocation system, MM assigned a rank to each index within each measure and then gave a final overall rank to each university based on the final ordering of total points assigned to all of the indices over all measures.

Mt. Allison was ranked first overall in Undergraduate universities, with University of Quebec (UQ) at Hull last; McGill was first in Medical/Doctoral universities, with Manitoba last; Simon Fraser was first in Comprehensive universities, with UQ/Trois Rivières last.

The comments to follow describe several difficulties in interpretation of these data. Although the analyses referred to below were investigative and somewhat exploratory, they examined the general hypothesis that empirical correlations between different component parts of the MM data would be consistent with the ranking results, and with the manner in which MM conceptualized and portrayed its overall findings.

## Macleans' Road Map

Due to the conceptual and actuarial omission of evaluative data concerning local social/demographic characteristics, overall missions, philosophies, and programs – including many which are unique to a given university – it is difficult to compare, contrast, or reconcile much of the ranking data.

The data allow no means by which one may subjectively weight or reliably discriminate between the various measures themselves or between their component indices, particularly when these parameters themselves turn out to be inconsistently related. St. Thomas, which MM refers to as having "the intimacy of a small institution" with only 1,733 full time students, still ranks third,

eleventh, tenth, and thirteenth (higher ranks meaning worse ratings) on the four indices of the Class Size measure. Mt. Allison is placed first overall among Undergraduate universities, yet stands 15th on two indices of the Classes measure (see above), 13th in library acquisitions, as well as seventh, eighth, and ninth on three other indices. UQ/Trois-Rivières was ranked last in Comprehensive universities, yet is ranked first in financial support from its alumni and fourth in two indices of Classes. Do alumni from Trois Rivières feel thirteen times better, or better at all, about their university than do alumni from Regina? While the question itself is of course absurd, its type is not totally out of line with a straightfaced approach to the ranking data and the general idiom in which they tend to be interpreted. In MM the data indeed are tabulated to show the “winners” at the top and, of course, the losers at the bottom.

How also might one interpret the size and significance of rank changes over time? Carleton, for example, was ranked nearly last (44th) in the 1991 MM rankings of all universities, yet (although the change may be due partly to changes in Carleton’s method of submission of data to MM) placed sixth in the 1992 rankings for Comprehensive universities. In only 12 cases (23%) in the 1993 rankings did a university receive the same rank from MM as it obtained in 1992. It should be noted that MM imposed 50-point penalties upon Carleton and Memorial for declining participation in the 1993 MM rankings. In its article, MM refers to these universities as “dropouts” but included data, taken mainly from the previous year, with which to include them in the current rankings.

Although it correctly states that no one parameter can unduly affect the rankings, MM did report the criticisms of Memorial president Arthur May, concerning validity of the criteria used, that the ranking procedures themselves are ultimately subjective and flawed. (For example, in the sense of criterion validity, to what extent do alumni contributions measure reputation?) The University of Calgary’s Vice President-Academic Joy Calkin also labelled the MM rankings accurate but “irrelevant” since they do not bear on the unique mission of each university. MM does not comment on, or adjust in its procedures, for Carleton president Robin Farquhar’s observation that Carleton’s acceptance of students with lower entrance grades penalizes that university in a ranking system which gives points for higher entrance grades. Such an admissions philosophy, which might as easily be praised as an example of democratization and increased accessibility to higher education, also applies to universities other than Carleton, for example, Laurentian, Lakehead, Windsor, Manitoba, and many others. MM also disregards similar comments from Brandon University; namely, that the rankings are insensitive to each university’s unique individual strengths. MM does remind readers that Brandon, in 1992, finished 13th out of 18 universities in its classification.

MM seems to regard academic criticisms of its procedures, such as those from Memorial and Carleton, as weak (“scholarly hairsplitting”), as opposed to noncritical comments from other academics (or CEOs). It cites without argument the views of one executive who portrays academics as afraid of self-evaluation and likely to find fault with any method of evaluating higher education, whatever it might be. Information about the job market is included as a major focus for evaluating universities, including unemployment rates for 14 different academic programs. Interestingly, this stands in contrast to the tone of what one of MM own authors, Ann Johnston, claims (p. 29) a university should be chiefly providing for students “a chance to be heard, and to learn to debate, analyze, and think.” Unfortunately, there seems to be no way of demonstrating that the realization of such goals is isomorphic with a university’s being of higher rank.

### **Pure Gold**

MM informs students they will discover “pure gold” in the rankings and various details provided about Canada’s universities. It indicates that students choosing a university will be guided mainly by three criteria: Class Contact (“a premium on small classes. . .”), Research (the “most vibrant and respected. . .”), and another termed Value Added (“Who improves their students the most?. . .”). The Value Added criterion implies here that MM believes students might well select a university with the sincere expectation that it will “raise” their grades more than will another. Concerning these three criteria, one observes that UQ/Montreal ranks first on Value Added, yet ranks sixth overall among Comprehensive universities and much lower on several indices which comprise the six main evaluative measures. Manitoba ranks second in Value Added, yet is last overall among Medical/Doctoral universities. Laurentian ranks highly (third) in Value Added, yet it comes 19th among Undergraduate universities. Acadia places second in Class Contact, yet is not even on the list of the top fifteen universities in either Value Added or Research. Exactly the same is true of Trent which was placed first in Class Contact. Space limitations allow for identification of only a portion of the anomalies in the MM students’ “road map.” It is interesting that MM (p. 36), despite its verbal acknowledgement of great differences in their mission, size, history, and geography, states that McGill is like UQ/Montreal, New Brunswick is like Manitoba, St. Francis Xavier is like York, and Queen’s is like Saskatchewan. These pairs of schools, for MM, are described as resembling twins separated at birth.

Further complicating the situation for students (MM’s client-consumers) is that, in many cases, the indices comprising each of the six main measures (see above) are themselves unrelated empirically and/or conceptually. Space limitations allow

listing here only a very small portion of these. For example, for Medical/Doctoral universities, Spearman rho rank correlations ( $N = 15$ ; with alpha at .05, one-tailed) based on the MM published rank data for these indices, show that alumni support and results from the MM "reputational survey" are uncorrelated. Also uncorrelated, using the same criteria, are library holdings and acquisitions, proportion of students graduating and students' incoming grades, as well as proportion of faculty with doctoral degrees and proportion of faculty winning awards. For Undergraduate universities ( $N = 23$ ), library holdings and acquisitions are found to be unrelated, as are the proportion of students graduating and their mean entrance grades. For Comprehensive universities ( $N = 13$ ), alumni support and reputational survey results, library holdings and acquisitions, and proportion of doctoral faculty and proportion of full time faculty winning awards, are all unrelated.

Overall ranks for 1992 and 1993 were highly correlated, with Spearman's rho = .95 and .93 for Undergraduate and Medical/Doctoral universities respectively. The rho coefficient was somewhat lower for Comprehensive universities (rho = .77). While these three correlations are high, there remains some residual change in ranks over the span of one year. For example, about 41 per cent of variance, in the case of Comprehensive universities, is not totally related to or explained by the rank for the previous year.

MM published additional data based on 1990 Statistics Canada information in which graduated students evaluated their universities according to class size, quality of teaching, job preparation, facilities, faculty access and whether they would return to the same institution again. These data (p. 47) are further crosstabulated by type of university, province, and type of curriculum studied. The vast majority of the surveyed students indicated high satisfaction with their university experiences on these criteria. By inspection, it is clear that there are few if any significant differences in any of these data. MM, however, in the absence of guidelines as to interpretability or statistical significance, searched for the two criteria on which the range of responses was greatest and claimed that the highest versus the lowest values still represented substantive differences. This, unfortunately, confounds sameness with equality. That is, universities with different point totals are construed as being different in rank, "greater than," "less than," and so on. It does not recognize the possibility, given the properties of rank data (see below), that they could differ in these totals yet show no functional difference in real terms, or that they could have the same point totals yet conceivably be different in terms of their value or attractiveness to students.

### Interpretation of Ranks

Although MM used a point allocation procedure to assess the six main measures and their component indices, its final published data are presented not in point totals but in the form of ordinal, that is, rank data. While rank data can be informative to some degree, differences in ranks are not amenable to meaningful comparative interpretation, either in a general sense or within any particular range of ranks. The properties of an ordinal scale are not isomorphic to the system of numerical analysis known as arithmetic (Siegel, 1959). Interpretation of differences between ordinal ranks is thus problematic even when the underlying scaled variable is simple, noncontentious, and linear (such as height or weight). It is vastly more difficult when such a variable is complex, contentious, and nonlinear. Moreover, in the present case, if there are “real” differences between certain pairs of universities, but not between other pairs, not an unreasonable possibility, the result is that the rank data have then only the properties of a nominal (that is, classificatory) and not even of an ordinal scale. In the present perspective, while academics may be conversant with the limitations of ordinal or nominal data, many readers of MM, among them a large percentage of students, parents, and members of the media, will likely not be and will, therefore, be prone to making fallacious comparisons, contrasts, or other misinterpretations of the data.

There are no clear guidelines for conceptualizing or measuring apparent differences both within and between the six main measures of universities or between the component indices of each. Class size, as a single example, has greater or lesser import depending on the type of course involved and probably also on the characteristics of the students and professors therein. Should a student interested in psychology avoid attending the University of Toronto because he or she will experience large classes, at least in the early stages? Is the difference between a class of 500 and one of 30 the same as that between a class of 30 and one of five? Within what ranges are differences important and in what ranges are they insignificant? In terms of instructor “quality” it might be pointed out that some universities employ graduate student instructors, for example, in introductory psychology courses. These students, while they may be several years away from holding doctoral degrees, are frequently given highly positive course ratings by students. But even if the data were consistent and interpretable, what are the upper limits to how seriously and how literally a student could consider the MM rankings? Is a given student better off (indeed, five times?) being taught by a professor whose doctoral degree was from Simon Fraser compared to one whose degree was from York? How many and what

type of clear differences, for example, should a student perceive as existing in a choice between New Brunswick (ranked eighth in Comprehensive universities) and Simon Fraser (ranked first)? Such questions are essentially rhetorical and cannot be answered with data of the type published by MM.

On the assumption, for now, that the MM point totals (and resulting ranks) are in some sense meaningful at least as ordinal data, one can examine to what extent lower-ranking universities differ from higher-ranking ones, that is, in terms of ranks on the six main measures used as evaluative criteria by MM.

Without advancing explicit hypotheses or predictions, the published rank data from the top and bottom subgroups (halves) of the universities within each of the three categories identified by MM were thus explored using Mann-Whitney U-tests. These tests examine the significance of differences in data, in the form of ranks, taken from independent samples of subjects (universities). Spearman rho (rank-based) correlations between universities' overall MM rankings and each of the six MM measures, setting alpha at .05, were also computed. In these analyses, each university's score (rank) on the six measures was the mean of the ranks given to its component indices by MM (pp. 30-35).

For Undergraduate universities, the U-tests, with alpha at .05, one-tailed, showed that the top ( $\underline{n} = 12$ ) and bottom ( $\underline{n} = 11$ ) subgroups did not differ significantly in terms of mean ranks for Faculty, Finances, or Library, although all six measures showed significant Spearman rho correlations with overall ranking. An exploratory multiple regression (discriminant) analysis, with subgroup membership (top vs. bottom) as the dependent variable, showed that while the six measures together showed the expected significant regression effect ( $F = 3.82, p < .01$ ), none of the six measures was a significant individual predictor independently of the joint effects of the remaining five.

For Medical/Doctoral universities, the U-tests showed that the top ( $\underline{n} = 7$ ) and bottom ( $\underline{n} = 8$ ) subgroups did not differ significantly in Classes, Finances, or Library. The Spearman rho correlations between overall rank and the six measures were also nonsignificant for these measures. An exploratory multiple regression analysis, with subgroup membership as the dependent variable, showed that while the six measures together showed the significant regression effect, ( $F = 6.92, p < .007$ ), only Faculty and Reputation were significant individual predictors.

For Comprehensive universities, although the Spearman rho correlations were significant for Student Body and Finances, none of the U-tests was significant for any of the six measures, nor was the exploratory multiple regression  $F$  (or that of any of the individual predictors) significant in terms of discriminating between the top ( $\underline{n} = 7$ ) and bottom ( $\underline{n} = 6$ ) university subgroups.



In examining the Pearson  $r$  intercorrelations (which yielded results comparable to those using Spearman's rho) between the six measures, using mean ranks as scores, the number of significant ( $p .05$ ) intercorrelations was relatively modest. For Undergraduate universities, Student Body was significantly correlated with Faculty, Finances, Library, and Reputation. Finances was significantly correlated with Library.

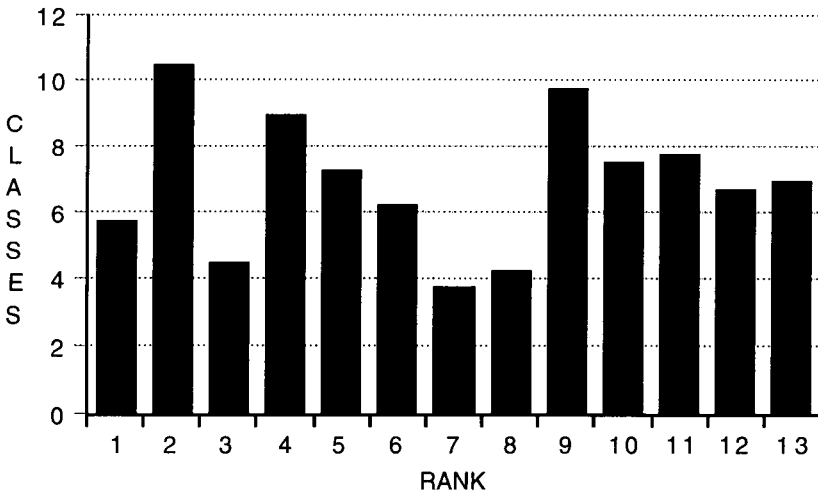
For Medical/Doctoral universities, Library was significantly correlated (negatively) with Student Body, that is, more favourable class size was related to less favourable library holdings and acquisitions. Also, Student Body was significantly correlated with Faculty.

For Comprehensive universities, Reputation was significantly correlated with both Faculty and Classes, the latter in a negative direction. Library was significantly correlated with Finances, as was Student Body. Faculty was significantly correlated (negatively) with Classes.

Although space limitations allow mention of only one brief example, it is observed that the universities' overall rankings are in many cases related only erratically to their mean ranks on the six main measures used by MM. A single example appears in Figure 1 in which overall rank is plotted against the Classes measure for Comprehensive universities. The figure shows further that the two

---

Figure 1  
Overall Rank Plotted Against Mean of Ranks for Classes Measure



Comprehensive Universities

---

best universities actually come seventh and eighth in the rankings and that the second best university stands worst of all in terms of its mean rank.

In addition, Spearman rho correlations were computed between the universities' overall rank and four selected, specific indices, that is, ranks for percentage of faculty with Ph.Ds, first and second year class size, operating budget, and proportion of students graduating. With a  $p$  level of  $=.05$ , one-tailed, as a criterion, it is observed that, for Medical/Doctoral universities, only percentage of faculty with Ph.Ds was significantly related to overall rank ( $\rho = .64$ ,  $p < .01$ ). For Comprehensive universities, only proportion of students graduating was significantly related to overall rank ( $\rho = .80$ ,  $p < .001$ ). Notably, for Undergraduate universities, none of these specific indices was related significantly to overall rank.

There may also be a self-fulfilling and reciprocal relationship between the ranking procedures and the universities' general popularity. That is, those universities mentioned most frequently may well be the most likely to benefit from subjective judgements about such measures as "reputation," "leaders of tomorrow," and so on, and most likely to influence the stereotyped images and perceptions of prospective students. In their analysis of current events concerning academic matters, the media also show a consistent tendency to solicit spokespersons from well-known, "larger centres." In the three MM general commentary articles in its November 1993 issue, that is, those not focusing on specific organizations or on details of the ranking procedures, MM mentions most universities not more than once and some not at all, while one is mentioned ten times. In its commentaries about the outcomes and the three overall "winners" in the Comprehensive, Undergraduate, and Medical/Doctoral categories, in only one instance is a university other than the "winner" mentioned at all.

In summary, aside from issues surrounding the suitability of the main evaluative criteria, the main pitfalls of the MM ranking approach seem to involve the: omission of evaluative criteria concerning the goals of particular universities and programs, difficulty in making reliable and valid discriminations both between and within the main measures and their component indices, uncertainty in accounting for changes in rank over time, concentration on criteria not likely to be realistically usable or interpretable by students; problems in interpretation of ranked data; presence of unreliable differences and inconsistencies between lower and higher ranking universities; and, to some degree, the possibility of self fulfilling mechanisms in the general interpretation of the findings. Some concluding comments follow from a different perspective on the MM approach and its effects.

### **Practical Implications of A “Rankings” Approach In University Selection**

As described above, MM chose to define quality of a university in terms of certain measures and their component indices. Many aspects (though clearly not all) of the relationship between these indices and overall rank must therefore be true “by definition.” It is doubtful, in any case, that most students will be in a position to rationally “choose” their university using these parameters, especially in view of academic, financial, personal, and geographical constraints. That is, in considering the rankings’ portrayal of universities’ strengths and weaknesses, it is unclear how students might “know” whether to choose a university for one or more of its supposed strengths or, alternatively, to avoid it for one or more of its weaknesses.

Aside from the “hard data” of a rankings approach, students need realistically usable information about specific programs and about financial matters. Unique programs undertaking particularly to combat exclusivity and elitism, for example, by making higher education more accessible to mature students, or to increased numbers of women, minorities, or other groups, also seem to imply and require different types of potential evaluative criteria. Many specific programs have unique local relevance and impact within a university’s geographical area. Moreover, a university could well institute significant changes in the academic content of one or more of its programs (or suffer cutbacks); yet such changes would be highly unlikely, given MM’s choice of evaluative criteria, to cause such a university to change in overall rank.

Perhaps not all students are appropriately equipped or motivated to benefit maximally from Cardinal Newman’s vision of “the idea of a university.” In turn, many are faced with the problem of overcoming dysfunctional learning habits and attitudes about academic goals, nourished by a general climate of mediocrity throughout their previous educational careers. The ranking approach tends to further reinforce many students’ predispositions toward intellectual passivity and submissiveness by continually referring to their great need for help and educational guidance. This reference, in turn, implicitly discourages them even further from trusting their own initiatives, skills, and self-explorations, as well as their own instincts about themselves and their personal needs. This of course assumes that students do in fact constitute a realistic audience for the ranking approach in the first place. It is unsettling that, while MM finds that students and alumni generally show high regard for what they have received in the way of higher education, Bercuson, Bothwell, and Granatstein (1984) claim in *The Great Brain Robbery* that Canadian universities are on “the road to ruin.”

Lastly, the rankings approach is alienated from a significant side-effect of the published rankings, and of their national circulation, upon those being helped. This factor, which has been voiced frequently by undergraduate students, concerns students' personal perceptions and feelings, both about themselves and about "their" university. As put by a student in one of the author's classes, referring to one of the "lower" schools, "How are the students there going to feel now. . ." Another said "Now, after this, they're going to think they aren't as good as the others; they were only ranked half as high. . ." These sentiments appear superficial, even banal, but are real in their effects. Fueled by the consumerist perspective that universities can be ranked or "rated" not unlike toasters or VCRs, these effects are a factor in students' overall sense of security and possibly even their academic performance. Until significant improvements can be made which will allow valid inter-university comparisons and contrasts, the approach of future students to university selection should not rely too heavily or naively on the statistical contrivance of ranking procedures.

### References

- Bercuson, D., Bothwell, R., & Granatstein, J. (1984). *The great brain robbery*. Toronto: McClelland and Stewart.
- Maclean's*. (1993, November 13). A Measure of Excellence. Vol. 106(46).
- Siegel, S. (1959). *Nonparametric statistics*. New York: McGraw Hill.