

## Rankings of Canadian Universities, 1995: More Problems in Interpretation

---

**STEWART PAGE**

*University of Windsor*

### Abstract

A critical perspective is presented in regard to rankings of Canadian universities by *Maclean's* magazine, November 20, 1995. Present comments are based, in part, on a previous analysis (Page, 1995) of the *Maclean's* rankings from 1993 and 1994. Several pitfalls in the ranking procedures, and results of some analyses of the 1995 ranking data, are briefly outlined.

---

### Résumé

L'auteur présente son évaluation du classement des universités canadiennes établi par le magazine *Maclean's* dans le numéro du 20 novembre, 1995. Ses remarques se fondent en partie sur une analyse antérieure (Page, 1995) des classements effectués par ce magazine entre 1993 et 1994. On soulève plusieurs éléments dissimulés inhérents à la méthode de classement utilisée et les résultats de quelques analyses des données de classement pour l'année 1995.

---

Additional tabulations or details of the present analyses may be obtained upon request from Stewart Page, Ph. D., Professor and Head, Department of Psychology, University of Windsor, Windsor, Ont. N9B 3P4. The author would like to thank three anonymous reviewers of the *Canadian Journal of Higher Education* for comments which were helpful in revising the present paper.

In its November 20, 1995 issue, *Maclean's* magazine (MM) published its fifth annual rankings of Canadian universities. The expressed intention of this latest enterprise was to "take the measure of Canadian universities" (MM, p. 31) and to provide a "ranking road map" with which to judge universities' strengths and weaknesses.

The following presents a brief summary of pitfalls in the 1995 rankings and overall statistical approach taken by MM. It is hoped that greater attention to such pitfalls could expedite future debate on the wider issue of how university evaluation and accountability might be conceptualized and addressed in a public forum such as that provided by a mass circulation magazine. Regarding its procedures and methodology for 1995 (and 1994), MM (p. 31) published a more complete description of these compared to that given in 1993 and earlier. It did, however, retain its basic philosophy of summation and conversion of preliminary evaluative data (point totals) to ranks (first, second, etc.) and of constructing from these a linear "ranking" of universities.

### **Measures Used**

For 1995, MM again classified universities into Medical/Doctoral ( $N = 11$ ), Comprehensive ( $N = 9$ ), and Primarily Undergraduate ( $N = 19$ ) categories, according to its definitions concerning the extent of a university's involvement with graduate programs and research. According to MM (p. 31), the spirit of the 1995 ranking analysis was that "the universities in the three categories are treated as separate but equal."

As in 1994, questionnaire data were compiled according to six main Measures each comprised, in turn, of several indices. The following Measures were used: Student Body (defined by indices of student ability, such as the grade average of incoming students); Classes (indices of class size and quality, such as percentage of classes taught by tenured faculty); Faculty (indices of faculty members' academic level, rank, and grant record); Finances (indices of budget, student services, and awards); Library (indices assessing holdings and collections); and Reputation (indices based on frequency of alumni support and on a Reputational survey sent to senior university officials, high school guidance counsellors,

and chief executive officers of Canadian corporations). Summed over the six Measures, a total of 22 indices were used for Medical/Doctoral universities, 21 for Comprehensive universities and 20 for Primarily Undergraduate universities.

As before, MM determined for each university its rank on each index within each Measure and then assigned a final overall rank (see MM, 1995, pp. 24-29) to each university based on its comparative standing after summing ranks over all indices, over all Measures.

In 1995, Mt. Allison was ranked first overall in Primarily Undergraduate universities with University College of Cape Breton last; Toronto was first in Medical/Doctoral universities with Calgary last; Victoria was first in Comprehensive universities with Concordia last.

### **Pitfalls in MM's Ranking Procedures**

As previously described (Page, 1995; Page, in press), several pitfalls occur in interpretation of these data. These become evident when examining the presumption of internal consistency and validity, including the related notion that relationships between different parameters in the MM data should be generally consistent with the overall (final) ranking results. The following points apply to the MM ranking procedures for 1995:

1. The final MM data are presented as ordinal, that is, rank data. As described in Page (1995), differences in ranks are not amenable to meaningful comparative or mathematical interpretation either generally or within any particular range of ranks (Siegel, 1959). Interpretation of differences between ranks is thus problematic even when the ranked variable is simple, noncontentious, and linear (such as, for example, height or weight). Moreover, in the present case, if there are 'real' differences between certain pairs of universities but not between other pairs, the result is that the rank data then have then only the properties of a nominal (that is, classificatory) and not of even an ordinal scale. Most readers of MM will likely be unfamiliar with the limitations of rank data and therefore prone to fallacious comparisons, contrasts, or other misinterpretations. Such errors are further encouraged by practices such as

MM's concentration on an evolutionary metaphor which implies that lower ranks are those of the less fit.

In like spirit, MM utilizes a noncritical terminology which includes parameters such as *drawing power* or *graduation rates*, and which labels some universities as *winners* and others, by implication, as *losers*. Moreover, as with last year, MM further summarizes the 1995 ranking analysis with pop metaphors. Thus, in contrast to others, Toronto has *clear-cut thinking*, Victoria shows a *no-nonsense approach to learning*, and Mount Allison *embraces the new and innovative*.

2. Many of the indices comprising the six main Measures (a rank being available for each index) are unrelated to each other. Spearman rank-order ( $\rho$ ) correlations, which indicate the degree of (linear) association between ranks for two variables in a given sample, were computed for each possible pair of indices used by MM, that is, here pooling the indices over all Measures.

For Medical/Doctoral universities ( $N = 11$ ) the mean number of significant correlations between any single index and another, disregarding sign, was 3.59, using a significance criterion of  $p < .05$ . Of the total of 231 correlations between these pairs of indices, 79 (34%) were significant; approximately 12 (5%) could be expected to be significant by chance, at  $p < .05$  (see Author Notes).

For Comprehensive universities ( $N = 9$ ), the mean number of significant correlations between one index and another, disregarding sign, was 2.61. Of the total of 210 correlations between pairs of indices, 55 (26%) were significant, of which approximately 11 could be expected to be significant by chance.

For Primarily Undergraduate universities ( $N = 19$ ), the mean number of significant correlations between one index and another, disregarding sign, was 3.75. Of the total of 190 correlations between pairs of indices, 75 (39%) were significant of which approximately ten could be expected to be significant by chance.

Spearman  $\rho$  correlations were also examined between all possible pairs of indices, within each Measure.

For Medical/Doctoral universities, nine of a total of 33 within-Measure correlations (27%) were significant at  $p < .05$ ; none was significant within the Classes or Reputation Measures.

For Comprehensive universities, nine of a total of 27 within-Measure correlations (33%) were significant. None was significant within the Library or Reputation Measures.

For Primarily Undergraduate universities, six of a total of 25 within-Measure correlations (24%) were significant. None at all was significant within the Finance, Library, or Reputation Measures.

3. Many of the indices are unrelated to MM's final rankings. Spearman  $\rho$  correlations were computed between each university's final rank, as assigned by MM, and its rank on each of the indices comprising the six main Measures.

For Medical/Doctoral universities, considering 22 indices, ten such correlations (45%) were significant at  $p < .05$ . The mean  $\rho$  for these was .71.

For Comprehensive universities, considering 21 indices, eight correlations (38%) were significant. The mean  $\rho$  for these was .77.

For Primarily Undergraduate universities, considering 20 indices, 11 correlations (55%) were significant. The mean  $\rho$  for these was .61.

4. The universities' mean ranks on the six main Measures are not uniformly or strongly related to overall ranking. Spearman  $\rho$  correlations between overall MM ranks and mean ranks on each of the six MM Measures were computed; in these analyses then, a university's score (rank) on each Measure was the mean of the ranks given by MM to its component indices.

For Medical/Doctoral universities, of the six Measures, the mean ranks for Reputation ( $\rho = .85$ ), Student Body ( $\rho = .82$ ), and Faculty characteristics ( $\rho = .85$ ) were significantly correlated with overall ranking at  $p < .05$ .

For Comprehensive universities only one such correlation, that concerning mean ranks for Student Body ( $\rho = .86$ ), was significant.

For Primarily Undergraduate universities, five of the six correlations were significant, that is, concerning mean ranks for Finances ( $\rho = .56$ ), Library resources ( $\rho = .51$ ), Student Body ( $\rho = .85$ ), Reputation ( $\rho = .75$ ), and Faculty ( $\rho = .64$ ).

5. Mean ranks on the six Measures are not strongly related to each other or to the Measures' various component indices. The matrix of

Spearman  $\rho$  intercorrelations was computed for the universities' mean ranks on each of the six Measures (as defined in the preceding section). Spearman  $\rho$ s were also computed between mean ranks on each Measure and ranks for all indices used by MM.

For Medical/Doctoral universities, of a total of 147 correlations comprised of both types just described, 42 (28%) were significant. The mean number of significant correlations between a university's mean rank on one of the six Measures and another Measure's mean rank or rank on one of the 22 MM indices was 7.00. Mean ranks for Measures concerning Classes, Finances, and Library resources were not significantly related to mean rank for any other Measure. Mean ranks for Student Body, Faculty and Reputation were in each case significantly correlated with those for two other Measures.

For Comprehensive universities, of 141 such correlations, 28 (19%) were significant. The mean number of significant correlations between a university's mean rank on one of the six Measures and another Measure's mean rank or rank on one of the 21 MM indices was 4.66. Only the mean ranks for Finance and Faculty were significantly correlated with the mean rank on even one other Measure (each other's).

For Primarily Undergraduate universities, of 135 such correlations, 55 (40%) were significant. The mean number of significant correlations between a university's mean rank on one of the six Measures and another mean rank or rank on one of the 20 MM indices was 9.16. The mean rank for the Faculty Measure was unrelated to those for all other Measures. Mean ranks for Library and Reputation were significantly related to those for one other Measure each. Mean ranks for Classes and Finance were each related significantly to those for two other Measures each. Mean ranks for Student Body were significantly related to those for three of the remaining five Measures. It is thus noted that, in general, the various Measures and indices data for Primarily Undergraduate universities are less inconsistent, in comparative terms, with the final overall rankings of these universities than is the case with Medical/Doctoral or Comprehensive universities.

6. If it is maintained that the overall MM ranks are somehow meaningful at least as ordinal data, it would seem reasonable to explore to

what extent lower-ranking universities differ from higher-ranking ones, for example, in terms of their mean ranks on the six Measures and on the indices upon which the Measures are based. The published rank data from the top and bottom subgroups (halves) of the universities were thus explored using the Wilcoxon Rank Sum test (equivalent to the Mann-Whitney U-test). This test examines the significance of differences in ranked data on a specified parameter taken from two independent samples of subjects (universities). Universities were compared by assessing whether the rank scores of the top half, on all indices, and on the mean of ranks given to the indices for each of the six main Measures, were significantly different from those of the bottom half.

For Medical/Doctoral universities, the Wilcoxon tests showed that the top and bottom groups (halves) differed significantly, at  $p < .05$ , on six (27%) of the tests computed for the 22 individual indices. These included the two indices concerning Reputation, three concerning Student Body, and one concerning Faculty characteristics. The Wilcoxon tests also showed that, considering each university's mean rank on the indices comprising each of the six Measures, the top and bottom groups were significantly different on three of the six Measures. These were: Student Body, Faculty, and Reputation. An exploratory stepwise discriminant function analysis, using the SAS (Statistical Analysis System) STEPDISK procedure, was also computed. Considering the universities' mean ranks on the six Measures' indices as potential predictors of group membership, only one such predictor was retained in a single significant (at  $p < .05$ ) discriminant function. This was the mean rank of the indices comprising the Reputation Measure.

For Comprehensive universities, the Wilcoxon tests showed that the top and bottom groups differed significantly on seven (31%) of the 21 individual indices. These included two indices concerning Student Body, two concerning Faculty, two concerning Reputation and one concerning Library resources. Considering each university's mean rank on the indices comprising each of the six Measures, the top and bottom groups were significantly different on two of the six Measures, namely, Faculty and Finances. A stepwise discriminant function analysis, of the type described above, showed that only two predictors were retained in a single significant discriminant

function. These were the mean ranks of the indices comprising the Faculty and Student Body Measures.

For Primarily Undergraduate universities, the Wilcoxon tests showed that the top and bottom groups differed significantly on seven (35%) of the 20 individual indices. These included three indices concerning Student Body, three concerning Faculty and one concerning Reputation. Considering each university's mean rank on the indices comprising each of the six Measures, the top and bottom groups were significantly different on three Measures, namely, Student Body, Faculty, and Reputation. A stepwise discriminant function analysis, as above, showed that three of the six potential predictors were retained in a single significant discriminant function. These were the mean ranks for the indices comprising the Reputation, Faculty, and Classes Measures.

Results using the nonparametric Wilcoxon Rank Sum test, reported herein, were compared in each case to those using Kuiper's test and the Kolmogorov-Smirnov test, which also test whether two independent samples have significantly different ranks. Results from the latter two tests yielded in every case the same p levels as were generated by the Wilcoxon test.

7. MM (p. 41) offers readers a 'worksheet' which lists all indices used for the 1995 rankings. The worksheet invites students, 'after reading the charts,' to 'customize a shortlist' of universities. Yet, following the above results, the data allow no means by which students may weight or discriminate reliably between the Measures or between their component indices (referred to by MM as indicators). Moreover, these parameters themselves are not clearly related, conceptually and/or empirically, to each other and/or to overall ranking.

The worksheet also omits factors, such as geographical location and its correlates, as well as personal factors and other types of information which are typically involved in one's choice of universities. In the author's experience, students frequently indicate that their degree of choice is limited to a very few alternatives, if in fact a realistic choice exists at all. In practical terms then, the worksheet exercise is thus rendered rather unhelpful, since it turns out that many of the Measures and/or their component indices play little or no empirical or conceptual



role in analysis of what university a 'student' should attend. At the least, it is clear that many indices may (arguably) make a conceptual but do not make an empirical contribution to a given university's final placement within the overall rankings. MM also includes no evaluative data concerning local social/demographic characteristics, overall missions, philosophies, and programs of universities, including many which are unique to a given one. (Are Concordia and Cape Breton, respectively, really "separate but equal" to Waterloo and Mt. Allison?) Accordingly, it becomes even more difficult for students or others to compare, contrast, or reconcile much of the ranking data.

8. Inconsistencies and anomalies again occur in the data (Page, 1995), and again raise the issue of how seriously students should attempt to synthesize the indices and Measures. Although space limitations prohibit a complete listing of these, it is noted, for example, that Ryerson is ranked first in the "Leaders of Tomorrow" and "Most Innovative" categories, (that is, as rated by high school guidance counsellors, corporate CEOs, and academic administrators), yet is not among the top five in the "Highest Quality" category. Ryerson is also ranked second in the "Best Overall" category of the reputational survey, yet ranks seventeenth in overall rankings for Primarily Undergraduate universities. Mount Allison is ranked fourth in "Best Overall," yet is not ranked among the top five in the reputational "Most Innovative" and "Leaders of Tomorrow" categories. It ranks 13th and 16th on two of the evaluative indices, yet is ranked first in the overall rankings for Primarily Undergraduate universities. Remarkably, Waterloo has now been ranked first in all four of the reputational indices, for 1993, 1994, and 1995, yet is still ranked third in the overall 1995 rankings. Unfortunately, MM has to date provided relatively little information about its "reputational" surveys and their component indices. It would be interesting to know the degree of concordance between specific universities attended by CEOs and the CEOs' rankings of "best" universities, as well as the rationale by which universities should be accountable to the criteria or "indicators" of corporate success.

It is noted, now considering the data for the three university types pooled together, that 11 of the Spearman rho intercorrelations between

indices were negative in direction; that is, in such cases, generally lower ranks for one index were related to generally higher ranks for another. For Comprehensive universities, the Student Body indices for proportion of out-of-province students, and for proportion of students graduating, were correlated .96 in a negative direction. Similarly, for these universities, three of the rho correlations between the mean ranks on the six Measures and the various indices, were negative.

Straightfaced interpretation of the overall MM rankings is also subject to the effects of inevitable yearly fluctuations in many aspects of the underlying data, and, for that matter, in the number of universities "participating" in the ranking exercise. In the 1994 rankings, for example, Windsor ranked fifth on the index assessing proportion of students who graduate, yet, due to other fluctuations and because the proportion shifted by eight per cent over the ensuing year, it ranks first on this index in the 1995 rankings. It is interesting that, for 1995, Waterloo did not provide information to MM concerning the index for percentage of students from out-of-province. As it has done in some past instances as well (Page, 1995), instead of omitting data not provided, MM elected to impose a "penalty," in this case by assigning Waterloo to last place for this index in the Comprehensive universities category.

### Summary

Interpretation of the 1995 rankings of universities by MM is thus again subject to numerous pitfalls in interpretation. As described above, these include: uncertainty in interpretation of changes in particular ranks over time, lack of differentiation of universities according to the six Measures and component indices used by MM, and vulnerability to known problems in making valid and reliable comparisons or interpretations using ranked data. While in some evaluative situations rank data may be informative, MM again portrays the evaluation of universities as an impersonal, objective process of measurement, similar to *Consumer Reports'* publication of ratings for light bulbs or VCRs. As such, MM does not consider meaningfully the role of information outside its own Measures or indices, nor that of personal and subjective factors inherent in what-

ever comparative appraisal of universities that students may be inclined to carry out. These factors are, however, well illustrated in Pesaro's (1993, p. 5) astute comment to prospective students that "there is no best university, but there is a best university for you." In MM's continuing dalliance with "measuring" the sites of higher education, perhaps the wider and more realistic perspective implied in Pesaro's comment could be included as a component part of future analyses. If nothing else, such might help to render MM's ostensibly rational and data-driven conclusions more effective, more complete and, in effect, more humane. ❁

## References

- Maclean's* (1993). A measure of excellence. November 15, 1993, Vol. 106, no. 46.
- Maclean's* (1995). Universities: Measuring excellence. November 20, 1995, Vol. 108, no. 47.
- Page, S. (1995). Rankings of Canadian universities: Pitfalls in interpretation. *Canadian Journal of Higher Education*, XXV, 18-30.
- Page, S. (1995, in press). Counselling for humane values in a competitive world. *Guidance and Counselling*.
- Pesaro, J. (1993). The best university. *INFO*, 17. Toronto: Ontario University Registrar's Association.
- Siegel, S. (1959). *Nonparametric statistics*. New York: McGraw-Hill.