

The Sophistry of University Rankings: Implications for Learning and Student Welfare

Stewart Page & Kenneth M. Cramer
University of Windsor

Laura Page
Trent University

We present a data-based perspective concerning the recent Maclean's magazine rankings of Canadian universities, including cluster and other analyses of the 2007 and 2008 data. Canadian universities empirically resemble and relate to each other in a manner different from their formal classification and final rank ordering in the Maclean's system. Several pitfalls in ranking procedures, related to invalid and unreliable relationships among indices underlying the final ranks, are outlined, along with relevant findings from previous studies. In their present format, although they have become increasingly publicized and promoted, data based on the Maclean's system are of limited practical use to students. Perhaps more important, ranking exercises have unintended though potentially serious consequences in terms of the intellectual and personal well-being of students.

Introduction

In November 2007 and December 2008, *Maclean's* published its 18th and 19th annual rankings of Canadian universities. Indeed, the ranking of universities has become a popular exercise with which to assess and promote higher education in North America (e.g., Bruneau & Savage, 2001; Cramer & Page, 2007; Page & Cramer, 2002, 2003; Page, Cramer, & Page, 2008; Provan & Abercromby, 2000). The rank-

ing approach is similar to that used by publications such as *Consumer Reports*, in which goods or services are assigned scores, and then assigned relative rank standings.

In this paper, we summarize our examination of the 2007 and 2008 ranking data, and present implications of ranking exercises for the issue of student welfare.

Ranking Procedures

For this exercise, *Maclean's* classifies universities ($N = 47$) into three types: Medical/Doctoral universities ($N = 15$), which contain medical schools and have large graduate departments; Comprehensive universities ($N = 11$), which have no medical schools but may have graduate programs, and Primarily Undergraduate universities ($N = 21$).

The *Maclean's* ranking data have always reflected six main measures: Student Body (including indices of student ability); Classes (including indices of class size and percentage of classes taught by tenured faculty); Faculty (indices of faculty members' academic characteristics); Finance (indices of budget parameters and student services); Library (indices assessing holdings); and Reputation (indices based on alumni support and a reputational survey). Pooled over the six measures, *Maclean's* has typically used a total of 24 indices for Medical/Doctoral universities, 23 for Comprehensive universities, and 22 for Primarily Undergraduate universities. For 2007 and 2008, however, rankings were reduced to 14, 13, and 13 indices respectively, though were still grouped within the above six measures. Details about each index are available in every *Maclean's* annual university rankings issue.

Maclean's compiles preliminary data for each index separately, computes a weighted sum of points, then constructs an overall rank, with standings reported separately within each university type. In the 2007 data, McGill ranked highest in the Medical/Doctoral category. For Comprehensive schools, the University of Victoria ranked highest. For Undergraduate schools, Acadia and Mount Allison tied for highest. For 2008, McGill, Simon Fraser, and Mount Allison ranked highest in these categories, respectively.

In previous studies, we have examined the *Maclean's* ranking parameters and published data, including the extent to which the indices were intercorrelated and related to the overall final rankings assigned. We also have examined the degree to which higher-ranking universities were reliably different from those of lower rank. We have found consistently that the main measures and component indices

used by *Maclean's* are internally inconsistent and not reliably related either to each other or to final ranks.

Observations: 2007 and 2008 Ranking Data

For 2007, using Spearman ρ (rank-based) correlations, we find that many indices are unrelated to final rankings. The proportion of indices significantly related (i.e., at $p < .05$) to final rank, collapsing over the three university types, was .471.

For Medical/Doctoral universities, 8/14 indices (57.12 %) were related significantly to final rank. For Comprehensive universities, 4/13 (30.76 %) were related significantly to final rank. For Undergraduate universities, 7/13 (53.84 %) were related significantly to final rank. For each university type, as in previous studies, many of the ρ correlations were *negative*, that is, with higher final rankings correlated with lower rankings on certain indices, and vice versa.

Although they are conceptually similar across the three university types, the specific *Maclean's* indices in 2007 and 2008 empirically correlate weakly with each other, and also elicit different patterns of interrelationships and interpretations, depending on the university category under consideration.

For 2008, using Spearman ρ (rank-based) correlations, we find again that many indices are unrelated to final rankings. The proportion of indices significantly related (i.e., at $p < .05$) to final rank (11 of 14), collapsing over the three university types ($N = 47$) was .712. We note that relative rankings, as well as correlations between rankings and reputation, continue to change dynamically each year. Moreover, due to minor variations annually in the underlying data over time, the top several schools in each category necessarily shift back and forth in terms of final rank standings, relative to each other. For Undergraduate schools, the ρ correlation between the 2006 and 2007 rankings was .655, indicating that most of the 2007 rankings were not predictable directly from those of the previous year. For 2008, this correlation was higher, i.e., .942 ($p < .0001$).

For Medical/Doctoral universities in 2008,

8/14 indices (57.12 %) were related significantly to final rank. For Comprehensive universities, 4/13 (30.76 %) were related significantly to final rank. For Undergraduate universities, 5/13 (38.46 %) were related significantly to final rank. For each university type, as in previous studies, many of the *rho* correlations again were negative, that is, with higher final rankings correlated with lower rankings on several indices.

Higher Versus Lower-Ranking Universities

We assessed to what extent lower-ranking universities in 2007 differed from higher-ranking ones, in terms of the various indices in the *Maclean's* system. The top and bottom subgroups (halves) of the universities, within each type, were therefore compared using the Wilcoxon Rank Sum (Mann-Whitney *U*) test. The Wilcoxon test examines the significance of differences in ranked data on a specified index, taken from two independent samples (universities). For all universities pooled together, 12 comparisons of these indices were significant.

For Medical/Doctoral universities, the top and bottom groups (halves) differed significantly, at $p < .05$, on 6/14 individual indices (42.85%). For Comprehensive universities, the top versus bottom halves differed significantly on 3/13 (23.07%) of these indices. For Undergraduate universities, the top versus bottom halves differed on 3/13 (23.07%) of the indices. Thus, collapsing over the three university types, the top and bottom halves did not differ significantly in average rank on 28 of the 40 individual comparisons. For these comparisons, higher-ranking universities therefore had little or no difference from those of lower rank.

We note in passing that conventional mathematical operations, quantitative comparisons, or interpretations of differences, i.e., beyond observing the rank differences per se, are not possible with ordinal (i.e., rank-based) data (Siegel, 1959).

For the 2008 ranking data, the indices for the top and bottom subgroups (halves) of the universities, within each type, were again compared. For all

universities pooled together, 14 of these 40 comparisons (35 %) were significant.

For Medical/Doctoral universities, the top and bottom groups (halves) differed significantly, at $p < .05$, on 5 of 14 individual indices (35.71 %). For Comprehensive universities, the top versus bottom halves differed significantly on three of 13 (23.07 %), and for Undergraduate universities, the top versus bottom halves differed on six of 13 (46.15 %) of the indices. Thus, collapsing over the three university types, the top and bottom halves did not differ significantly in average rank on 26 of the 40 (65 %) individual comparisons. For these comparisons, higher-ranking universities therefore had little or no difference from those of lower rank.

Cluster Analysis of Universities

A vertical rank ordering of schools tends to exaggerate the apparent differences in final ranks, encourages over-interpretation of rank differences, and obscures similarities. From a different perspective, we therefore employed a cluster analysis (Landau & Leese, 2001) with which to examine patterns of interrelationship among the universities for the 2007 and 2008 rankings, across the three university types used by *Maclean's*. Using algorithms outlined by Ward (1963), squared Euclidian distances, as estimates of distance between schools, were calculated for the 47 schools, based on their raw scores for the 14 (13 for Comprehensive; 13 for Undergraduate universities) indices. The overall analysis for the 2007 rankings, including all universities, yielded five primary clusters. Unique clusters of schools were identified, in which the similarity of each member's corresponding profile was maximized, and intercorrelations among members were high. Schools were thus highly similar within a cluster, but dissimilar to those outside of the cluster.

Generally, members within each cluster showed considerable variation in their final rankings. In several cases, universities also belonged to clusters in a manner inconsistent with their membership in one of the three university categories. Thus, one may distinguish between formal a priori or conceptual similarity of universities within a particular type, as de-

fined by *Maclean's*, and empirical similarity as defined by quantitative analysis of the ranking indices themselves. In most cases, the pattern of relationships within and between clusters was not clearly reflective of rank differences between higher- and lower-standing universities within the three university types. In effect then, schools of different characteristics, programs, types, and rank standings may nevertheless be similar in terms of their scores on a particular set of indices.

For the 2008 data, the overall analysis, including all universities, again yielded five primary clusters. Generally, members within each cluster could be grouped into similar pairs or subclusters, and showed considerable variation in their final rankings and membership with reference to the three *Maclean's* university categories. In several cases, 'unlikely' pairs of schools turned out to be empirically similar. In effect then, schools of different characteristics, programs, missions, types, and rank standings may nevertheless show communality in their pattern of scores on a particular – arguably arbitrary – set of indices.

Discussion and Conclusions

Aside from the formal concerns of statistical data and comparisons, there remain other issues relevant to ranking exercises.

First, the ranking indices and specialized academic measures used by *Maclean's* are generally incongruent with the practical reasons for university attendance and selection typically reported by undergraduates themselves (e.g., Cramer & Page, 2007; Page, Cramer, & Page, 2008).

Second, the annual rankings have not generally reflected currently available studies of student satisfaction, wherein most students have indicated high levels of satisfaction with their own institutions, yet in which the highest ranking institutions, perhaps surprisingly, have often done relatively poorly. It is worth noting that, were a set of university rankings to be based on student satisfaction indices generally, it is likely that few if any significant differences in university-rank ordering would therefore be observed.

Third, an emphasis on rank standings ob-

scures the existence of individual programs and the unique missions of different schools, in the context of their particular location and type of student population.

Fourth, among many possible side-effects of ranking exercises, concerns their unintended effects upon the quality of a university's academic and intellectual spirit as they are experienced and perceived by students. These effects raise the possibility that the rankings may help to generate yet another form of the educational self-fulfilling prophecy at the post-secondary level (Blum, 1978; Page & Rosenthal, 1990; Rosenthal & Jacobson, 1968), indeed one which will affect some students adversely, and others positively. In very practical terms, students do not want to believe that their schools can be easily and publicly rank-ordered, no more than they believe that people should be ranked with indices defining them as haves or have-nots. Accordingly, the academic performance, general orientation, and perhaps overall world view, of those attending 'lower' universities continue to be threatened by exposure and reference to ranking exercises, and the typical interpretations which accompany them. Moreover, sizeable research literature now exists which has identified less effective thinking and less efficient intellectual performance in individuals who feel put down or stigmatized in some way. A recent series of studies has shown, for example, that being a member of a negatively stereotyped racial minority can adversely affect one's academic performance in situations in which the person believes that his or her intellectual ability is being assessed (Steele, 2004; Steele & Aronson, 1998; Kaplan & Saccuzzo, 2001). The notion that one's university ranking may be used to calibrate levels of 'smartness,' (as in the *Maclean's* question: "Where are the smart students?"), something like finding one's proper shoe size thus appears misguided and undoubtedly, we believe, harmful to many students. Such a perspective thus reinforces an educational idiom in which students are encouraged to attribute academic failure or attendance at lower-ranking schools to lack of ability, rather than to other possible explanations pertaining to behaviours under an individual's control (Dweck & Leggett, 1998). To this point, we note that several current research and

counseling programs are thus encouraging students to adopt more helpful and enlightened orientations to learning. The attribution of failure to lack of ability has been linked to many undesirable educational outcomes, such as decreased task persistence, as well as more apathetic and more passive or 'helpless' orientations to learning (e.g., Dweck & Leggett, 1998; Dweck & Molden, 2005).

Portrayals of smartness as a static and directly quantifiable ability are marketed within a culture modeled upon the metaphor of *Consumer Reports* – winners, losers, a 'best of all' school, and so on. Moreover, the rapidly expanding interest and literature on 'self-handicapping' (e.g., Urdan & Midgley, 2001), raises the researchable hypothesis for the future that students from lower-ranking schools and universities may engage in relatively higher levels of self-defeating and distracting activities, which in turn may serve effectively as excuses for lower academic performance. Future research must therefore be directed to question whether students from lower-ranking institutions suffer in terms of job-seeking activities and even whether, for example, their academic output is evaluated differently or less positively relative to the same input when shown by non-stigmatized students.

It is perhaps ironic that while Canadian schools have now (as of March, 2009) largely withdrawn their active support in terms of providing evaluative data upon demand to *Maclean's* magazine, several have nevertheless drawn upon their rank standings, strategically and selectively, with which to promote and market particular programs within the business of higher education. In view of the perceived inevitability of ranking exercises, and the resulting prospect of publicity, several schools naturally have engaged in lobbying efforts toward inclusion of additional ranking indices perceived as likely to enhance particularly their own relative rank standings.

We do not expect customers to evaluate (or reliably rank) all cars in a used car lot, to choose or agree upon the criteria to be used for such a task, or to possess the knowledge required to attempt meaningful comparisons. We similarly would believe that incoming students, for the most part, are hardly in a credible position with which to evaluate and meaningfully calibrate indices of the type which underlie

university ranking exercises. Moreover, such vulnerability is increased even more in the midst of unremitting political and lobbying activities on the part of university authorities, as well by as the siren call of advertising, employed deliberately and competitively by many schools, with which to attract student-customers. The student, with an individual and unique worm's eye view, is thus effectively a target, caught in the crossfire between the traditional ideals and presumed value of higher education, and the mundane reality of promotive marketing. In our opinion, and looking to the future, the characteristics of the underlying data, and the unintended but real consequences of annual ranking exercises, should now be of concern to parents, students, and authorities at all educational levels.

References

- Blum, J. (1978). *Pseudoscience and mental ability*. New York: Monthly Review Press.
- Bruneau, W. & Savage, D. (2001). *Counting out the scholars*. Halifax, NS: James Lorimer & Company.
- Cramer, K.M. & Page, S. (2007). Cluster analysis and rankings of Canadian universities: Misadventures with rank-based data and implications for the welfare of students. *Journal of Applied Multivariate Research*, 12, 183-198.
- Dweck, C. & Leggett, E. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273.
- Dweck, C. & Molden, D. (2005). Self theories: Their impact on competence motivation and acquisition. In A. Elliott & C. Dweck (Eds.), *Handbook of competence and motivation* (pp. 122-140). New York: Guilford Press.
- Kaplan, R. & Saccuzzo, R. (2001). *Psychological testing*. Belmont, CA: Wadsworth Publishing.

- Landau, S. & Leese, M. (2001). *Cluster analysis* (4th ed.). New York: Oxford University Press.
- Page, S. & Cramer, K.M. (2007). Ranking and evaluation of Canadian universities: Implications for student welfare. *Canadian Journal of School Psychology, 22*, 4-13.
- Page, S. & Cramer, K.M. (2002). Ranking of universities: Buyer beware. *Canadian Journal of Education, 25*, 297-310.
- Page, S. & Cramer, K.M. (2003). An update on the use of ranks in calibrating and marketing higher education. *Journal of Marketing for Higher Education, 13*, 87-101.
- Page, S., Cramer, K.M., & Page, L. (2008). *The sophistry of university rankings: Implications for student welfare*. Paper presented at Annual Meeting of the Society for Teaching and Learning in Higher Education, University of Windsor, Windsor, ON.
- Page, S. & Rosenthal, R. (1990). Sex and expectations of teachers and sex and race of students as determinants of teaching behavior and student performance. *Journal of School Psychology, 28*, 119-131.
- Provan, D. & Abercromby, K. (2000). *University league tables and rankings: A critical analysis*. Commonwealth Higher Education Management Service, Paper no. 30.
- Rosenthal, R. & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart, & Winston.
- Siegel, S. (1959). *Nonparametric statistics*. New York: McGraw Hill.
- Steele, C. (2004). *Social identity threat: How it affects intellectual performance, development, intergroup relations and what can be done about it*. Invited address at the Annual Meeting of the Jean Piaget Society, Toronto, Ontario.
- Steele, C. & Aronson, J. (1998). *Stereotype threat and the test performance of academically successful African Americans*. In E. C. Jencks, E. Phillips (Eds.), *The black-white test score gap* (pp. 401-427). Washington, DC: American Psychological Association.
- Urdu, T. & Midgley, C. (2001). Academic self-handicapping: What we know, what more there is to learn. *Educational Psychology Review, 13*, 115-138.
- Ward, J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*, 236-244.

Biographies

Stewart Page is a Professor in the Department of Psychology at University of Windsor, Ontario. His primary interests are in social psychology, evaluation of universities, personality theory and assessment, and issues in higher education.

Kenneth M. Cramer is a 3M Teaching Fellow and Professor in the Department of Psychology at the University of Windsor, Ontario. His interests are in educational psychology, the impact of university rankings on university students, and lecture engagement strategies.

Laura Page is Assistant Professor in the Department of Psychology at Trent University, Ontario. Her principal interests are in educational and adolescent psychology, including processes in identity formation and development of adolescents' beliefs and open-mindedness.