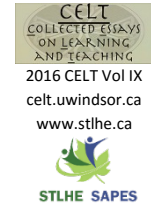# The Marketing of Canadian University Rankings: A Misadventure Now 24 Years Old

Kenneth M. Cramer, Stewart Page, Vanessa Burrows, Chastine Lamoureux, Sarah Mackay, Victoria Pedri, and Rebecca Pschibul
University of Windsor

*Based on analyses of Maclean's ranking data pertaining to Canadian universities published over the last 24 years, we present a summary of statistical findings of annual ranking exercises, as well as discussion about their current status and the effects upon student welfare. Some illustrative tables are also presented. Using correlational and cluster analyses, for each year, we have found largely nonsignificant, inconsistent, and uninterpretable relations between rank standings of universities and Maclean's main measures, as well as between rank standings and the many specific indices used to generate these standings. In our opinion, when assessed in terms of their empirical characteristics, the annual data show generally that this system of ranking is highly limited in terms of its practical or academic value to students. Among other difficulties with the interpretation of ranks, we also discuss the possibility that ranking exercises have unintended, though potentially serious, negative consequences for the intellectual and personal welfare of students.*

## Introduction

The exercise of rating or ranking life's various entities is commonplace today – from best weather and places to live, to restaurants and toasters, to recording artists and vacation spots. Forbes' top 500 companies and public institutions such as hospitals and even centres of higher learning are no exception (Aghaz, Hashemi, & Atashgah, 2015; Amsler & Bolsman, 2012; Huang, Chen, & Chien, 2015; Page & Cramer, 2001; Page, Cramer, & Page, 2008, 2010). Canadian publication *Maclean's* similarly aims to aid consumers in reaching a sound decision on where to attend college or university.[1] However, with increased public demand for external accountability, the transparency of these institutions becomes paramount (Allen & Bresciani, 2003; Shavelson & Huang, 2003), and spurs a spirit for additional (albeit valid) assessment data. As tools intended for monitoring institutional reputation and performance, Van Dyke (2005) suggests these report cards or "league sheets" as found in *Maclean's* are especially popular among students and parents, and are also becoming increasingly accepted in academia. Yet the exercise of ranking is not without its criticisms. They include (but are not limited to) the halo effect of reputation, arbitrary subjectivity, relative weightings of indices, lack of statistical rigor, and the limits of ranks as units of measure (Brooks, 2005; Clarke, 2002; Ferguson & Takane, 1989;

---

[1] Since their initial publication, Maclean's has instituted occasional minor adjustments to their university ranking procedures, although their main measures, component indices, and overall approach remain essentially unchanged.

Huang et al., 2015; Page et al., 2010; Provan & Abercromby, 2000; Siegel, 1959). As Salmi and Saroyan (2007, p. 52) note:

> *Notwithstanding their controversial nature and methodological shortcomings, university rankings have become widespread and are unlikely to disappear. Possible reactions, in the face of this rapidly expanding phenomenon, are to ignore, dismiss or boycott any form of ranking. Another, less extreme response is one that seeks to analyse and understand the significance and limitations of ranking exercises.*

We are presently pursuing the latter avenue, and sought to analyze the 2011-2015 data in the Canadian arena, as provided by *Maclean's.*

To address concerns surrounding institutional rankings as a global challenge, a UNESCO conference concluded with the cross-national mandate to evaluate and rank institutions of higher learning (Dill & Soo, 2005), which could direct educational policy through: (1) an international agreement on how to assess academic quality, (2) an evaluation of the impact such a ranking exercise might have on academic behaviour; and (3) an outline of salient public interests excluded from current institutional rankings. To be particularly useful though, Gormley and Weimer (1999) believed such an exercise should assess validity using measures that necessarily reflect valued social outcomes, but should also control for institutional differences among both students and relative resources. In short, they argue that league tables lack both the theoretical or empirical justification required for use of the selected measures.

Dill and Soo (2005) examined the institutional rankings across each of Australia, Canada, the United Kingdom, and the United States and concluded that typical institutional indices — divided into input, process, and outcome categories — had rendered relative consensus on input measures: incoming grades, study/faculty ratio, research grants, among others but little on both process and outcome initiatives. Moreover, a school's overall ranking was largely based on the amount of research conducted at the school (called the "American model"), which incidentally correlated negatively with student learning. Most noteworthy for the present study is their evaluation of the *Maclean's* Canadian ranking exercise, judged to be the most inadequate of the national systems reviewed, chiefly because they relied heavily on subjective rankings of reputation and utilized principally input measures. Previous research has specifically investigated the validity and interpretability of *Maclean's* rankings (Cramer & Page, 2007; Page et al., 2008, 2010), and similar conclusions were reached, namely that the indices selected by *Maclean's*: (1) did not perform adequately under the psychometric and statistical microscope, (2) were only somewhat relevant to the types of information sought by students and families in their choice of an institution of higher learning, and (3) may incite more harm than good concerning student welfare and institutional self-portrayal. We will similarly show how these outcomes remain unchanged in a five-year analysis of the most recent data.

Although on several occasions — either individually or en masse — Canadian schools have attempted to withdraw from *Maclean's* rankings (Salmi & Saroyan, 2007), the most recent data are now drawn largely from publicly available sources, such as Statistics Canada. Schools are divided into three categories: (1) Medical/Doctorate (with full medical training and a broad range of Ph.D. degrees), (2) Comprehensive (without medical training but still a sizeable range of graduate programs), and (3) Undergraduate (no medical training and limited graduate degrees). Further, six main measures are emphasized: (1) Student Body (e.g., students' past performance); (2) Classes (e.g., class size and percentage of classes taught by tenured faculty); (3) Faculty (e.g., faculty members' academic qualifications); (4) Finances (e.g., budget parameters and student services); (5) Library (e.g., acquisitions and holdings); and (6) Reputation (e.g., alumni support and reputational survey results). For 2015, the number of indices comprising each measure has been set at 12 within each category. Although the underlying component indices remain essentially as

before, the main measures have been renamed Students/Classes, Faculty, Resources, Student Support, Library, and Reputation. We will review the results of prior studies (1991-2010), and then proceed to show a comparable pattern for the 2011-2015 sets of data.

## Key Observations from Annual Data Analyses, 1991 to 2010

To illustrate our routine analysis plan, we presently outline the results from the 2010 ranking data (Page et al., 2010). To begin, a Spearman (rank-based) *rho* correlation analysis assesses the level of association between rank-based variables (viz. individual index rank against the final overall rank); we found that many indices were actually *unrelated* to final ranks. For each university type, as in all previous studies, many of the *rho* correlations were actually *negative* — where higher final ranks correlated with lower index values, and vice versa. For Medical/Doctoral universities, only 6 of the 14 (43%) possible correlations were statistically significant ($ps < .05$, replicated 19 times for every 20 investigations). For Comprehensive universities, 4 of 13 correlations (30%) were significant; and for Undergraduate universities, 5 of 13 correlations (38%) were significant. Although conceptually similar across the three *Maclean's* university types, inspection of the intercorrelation of indices for the 2010 data shows they correlate weakly and unpredictably with each other; that is, schools that rank highly on student bursaries may not be overall a highly ranked school. In practical terms, students and families likely lack the statistical acumen to properly analyze and interpret these data, as we have done presently.

We also assessed the extent to which lower-ranking universities differed from higher in terms of the *Maclean's* indices; herein we utilized the Wilcoxon Rank Sum test (Mann-Whitney U-test), which assesses the significance of differences in ranked data on a specified index, taken from two independent samples of universities. For all universities pooled together, only 9 of these 40 comparisons (22%) were significant ($p < .05$). For Medical/Doctoral universities, the top and bottom groups (halves) differed significantly on only 2 of the 14 (14%) indices; this was 3 of 13 (23%) for Comprehensive universities, and 4 of 13 (30%) for Undergraduate institutions. Thus, collapsing over the three university types, the top and bottom halves *did not* differ significantly in average rank on 78% individual comparisons, meaning that higher-ranking universities were little or no different from lower-ranking ones.

Finally, we employed Ward's cluster analysis (Landau & Leese, 2001) to examine interrelations and similarities among the universities for the 2010 rankings, across the three university types. This procedure identifies clusters or families of schools that are empirically similar via comparable index scores, and excludes those that are dissimilar. For each annual analysis, we have routinely found that the relations within and between clusters (i.e., groupings of empirically similar schools) are not clearly reflective of rank differences between higher and lower standing universities, or differences within or across the three university types. In several cases, unlikely groupings of schools are seen nevertheless to be empirically similar in terms of their pattern of scores on the indices contributing to their final ranks. In effect, schools of different characteristics, programs, missions, types, and rank standings may nevertheless show communality in their pattern of scores on a particular set of indices.

## Observations of Ranking Data, 2011 to 2015

As one notable change in 2011, *Maclean's* designated Brock University, Ryerson University, and Wilfrid Laurier University into the Comprehensive (rather than Undergraduate) category. However, our basic observations were highly similar to those from previous years. For all university types combined, the intercorrelations between specific indices were generally low, and only 23 of 40 (57%) possible *rho* correlations between indices and final rank were

significant (examples include student/faculty ratio, medical/research grants received, operating budget, and both student and library services). Furthermore, the Wilcoxon tests comparing higher versus lower ranked schools showed only 12 of 40 (30%) comparisons to be significant (with comparable variables from the previous analysis). Finally, a cluster analysis identified several clusters and sub-clusters, each containing family members whose coexistence (albeit improbable *a priori*) belonged to clusters empirically similar based on constituent indices. We thus found that schools of different types, or which appear dissimilar in other respects, may nevertheless turn out to be empirically similar in terms of their scores on an array of indices perceived to be worthwhile parameters of evaluation. Comparable findings, via the same statistical rigor implemented for the data for 2012 to 2015, indicated the same pattern of results as seen in previous years. To use the 2012 data as an example, 63% of *rho* correlations between specific indices and overall rank were significant for Medical/Doctoral universities, 75% for Comprehensive universities, and 38% for Undergraduate universities (see Table 1). For the Wilcoxon tests comparing, as before, the mean ranks of top versus bottom schools, 57% were significant Medical/Doctoral universities and 43% for both

Comprehensive and Undergraduate universities (see Table 2). The cluster analysis for the 2012 data yielded two primary and two sub-clusters, again with cluster membership largely unrelated to their members' (universities') general academic characteristics, overall rank standing, or university type. This pattern is consistent in the analysis of data from 2013 to 2015 inclusive. We are then left to conclude that, across these five additional years of data, little has changed with respect to sound statistical evidence to support the validity of *Maclean's* rankings.

## Discussion

Overall, the present analysis of *Maclean's* ranked indices from 2011-2015 corroborate those of prior studies (1991-2010), wherein (a) individual indices correlated with overall rank approximately half the time, (b) high versus low ranking schools were significantly different on roughly half the indices, and (c) cluster analysis produced largely meaningless and incomprehensible (albeit empirically similar) families of institutions (cf. Cramer & Page, 2007; Page & Cramer, 2001; Page, Cramer, & Page, 2008, 2010). That is, the ranking results generally illustrate

Table 1

*Percentage of Indices Correlating (p < .05) with Overall Rank*

| Year | Medical/Doctorate | Comprehensive | Undergraduate |
| --- | --- | --- | --- |
| 2011 | 57% | 43% | 43% |
| 2012 | 63% | 75% | 38% |
| 2013 | 43% | 50% | 46% |
| 2014 | 42% | 50% | 46% |
| 2015 | 47% | 42% | 42% |

Table 2

*Percentage of Indices Showing Significant Differences between Top- and Bottom-Ranked Schools*

| Year | Medical/Doctorate | Comprehensive | Undergraduate |
|------|-------------------|---------------|---------------|
| 2011 | 57% | 43% | 43% |
| 2012 | 57% | 43% | 43% |
| 2013 | 38% | 38% | 38% |
| 2014 | 38% | 38% | 38% |
| 2015 | 33% | 25% | 25% |

unreliability and interpretational difficulty in the various aspects or comparisons considered, regardless of university type or other parameter. Recent analyses in other countries have rendered the same difficulties (see Aghaz et al., 2015; Amsler & Bolsman, 2012; Huang et al., 2015). In different nations and over time, similar challenges emerge in the exercise of rankings, which we subsequently highlight, specifically: comprehensive indexing, relative index weightings, reputational subjectivity, institutional withdrawal, institutional rank manipulations, and, finally, negative student impact.

To begin, whereas we see the wider scope of components that *Maclean's* elects to include, we are nonetheless left to wonder how comprehensive the list is — that is, which key variables may be excluded. In particular, annual rankings typically do not reflect the results of available studies of student satisfaction (Brooks, 2005; Page et al., 2010). Students often indicate high levels of satisfaction and loyalty toward their own institutions regardless of their rank — where higher ranking institutions often perform relatively poorly on a given measure (Pike, 2004). This tendency is evidenced in the National Student Survey of Engagement (NSSE) data, which evaluate students' impressions of the strengths and weaknesses with respect to curriculum, instruction, and campus living. The nonsignificant relation suggests that student impressions of their educational experiences are largely independent from institutional characteristics.

Secondly, the final rankings of institutions of higher learning depend heavily on the relative weightings that data centres choose to assign to any number of indices. Stanford University sitting-president Casper (1996) even criticized the US ranking agencies for this practice. These weightings vary significantly across nations, rendering the UNESCO mission of cross-national comparisons rather arbitrary. As several have noted (Brooks, 2005; Provan & Abercromby, 2000), the rankings of institutions are inherently flawed by this embedded subjectivity.

A third issue concerns not just the role of institutional reputation in the calculation of overall ranks, but the rampant subjectivity uncovered in the methodology. Regarded as gossip and hearsay, critics argue that popularity contests of reputational data have a perpetuity seemingly immune to later adjustments to overall rank. This problem results from the high school principals and guidance counsellors, business CEOs, and other reputational experts being chiefly unfamiliar with the institutions they must evaluate. Whereas Dometrius, Hood, Shirkey, and Kidd (1998) suggested this institutional

unfamiliarity could be as high as 30%, raters would still provide their rankings (Brooks & Junn, 2002, qtd. in Brooks, 2005). One stark example of this subjective halo effect saw Princeton awarded the accolade of top ranked law school, despite not even having such a program (Brooks, 2005). Data out of US schools indicated reputation could be confidently predicted using just three variables: undergraduate selectivity, per-student expenditure, and the number of departments granting doctoral degrees (Astin, 1985, 1991). In short, *Maclean's* and other ranking agencies need to carefully evaluate the validity and impact of how reputation — an arguably salient but highly vulnerable element to students and families — factors into the overall ranking of an institution.

Furthermore, as early as 1993 — just 2 years after the ranking exercise was first introduced — institutions frustrated with either the process or results of *Maclean's* rankings have operated in a manner akin to Salmi and Saroyan's (2007, p. 52) advice: "Possible reactions, in the face of this rapidly expanding phenomenon, are to ignore, dismiss or boycott any form of ranking." In the Canadian context, both Memorial University of Newfoundland and Carleton University elected to no longer participate, as a protest to the methodology employed in the rankings exercise. Following a 1994 letter from McGill vice-chancellor Bernard Shapiro to *Maclean's* coordinating editor Ann Dowsett Johnston (Salmi & Saroyan, 2007), 15 universities elected to no longer participate. When the University of Toronto bowed out in 2005, *Maclean's* editors implemented freedom-of-information laws to obtain the data to compile rankings from those institutions who chose not to participate (Alphonso, 2006a, 2006b). The implication is that Canada's institutions of higher learning may no longer control the use and manipulation of their public data from ranking agencies such as *Maclean's*.

Far more worrisome still is the practice of institutional rank manipulations that may result from this exercise. That is, institutions set out to actively adjust their data to upwardly notch their overall rank. For example, it was uncovered that the University of British Columbia senior administrators urged faculty members to cap course enrollments in an effort to improve their position in *Maclean's* ranking system (Schmidt, 2004). Similar manipulations reported south of the border, involving US News Reputational Survey, implicated Cornell University, Clemson University, and the University of Florida (see Bastedo & Bowman, 2011; Lederman, 2009; Lee, 2009; Stevens, 2007).

Truly, though, it is the final matter of potentially negative student impact, as suggested by UNESCO (Dill & Soo, 2005), that makes us pause in our consideration of the overall value of the ranking exercise. We also offer this as a useful avenue for empirical research, since to date there exist no studies examining the relative impact of rankings (positive or negative) on student welfare. We may hypothesize that students from low-ranking schools will be made aware of publicized university rankings and their implied meaning about better students, better locations, and the implications for employment prospects; this may all pose a significant threat on more than just their personal identity and self-esteem, but also on the overall likelihood of their success. In view of our own analyses and in conjunction with other research cited, we would hypothesize that ranking systems, and their likely effects upon students' educational expectations, may well generate another form of the educational self-fulfilling prophecy (Rosenthal & Jacobson, 1968, 1992; Steele, 2004). We view future research on such a hypothesis as vitally necessary.

In conclusion, we urge readers — students, families, and the broader public — to demand that *Maclean's* provide evidence assessing the reliability and overall validity of its ranking system as it has evolved to date. We are privy to the need and use of, and even manipulation and abuse of, these data, and we likely will see the rankings of institutions of higher learning continue for decades into the future. However, our hope is for more responsible and accountable reporting of data in those years to come, to the doubtless benefit of our students.

# References

Aghaz, A., Hashemi, A., & Atashgah, M. S. S. (2015). Factors contributing to university image: The postgraduate students' point of view. *Journal of Marketing for Higher Education, 25,* 104-126.

Alphonso, C. (2006a, April 19). Schools snub Maclean's survey. *The Globe and Mail*, p. A8.

Alphonso, C. (2006b, August 15). Universities boycott Maclean's ranking. *The Globe and Mail*, p. A10.

Amsler, S. S., & Bolsman, C. (2012). University ranking as social exclusion. *British Journal of Sociology of Education, 33,* 283-301.

Allen, J., & Bresciani, M. J. (2003). Public institutions, public challenges: On the transparency of assessment results. *Change, 35*, 20-23.

Astin, A. W. (1985). *Achieving educational excellence: A critical assessment of priorities and practices in higher education*. San Francisco: Jossey-Bass Publishers.

Astin, A. W. (1991). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. New York: American Council on Education/Macmillan.

Bastedo, M. N., & Bowman, N. A. (2011). College rankings as an inter-organizational dependency: Establishing the foundation for strategic and institutional accounts. *Research in Higher Education 52,* 3-23. http://dx.doi.org/10.1007/s11162-010-9185-0 VIEW ITEM

Brooks, R. L. (2005). Measuring university quality. *Review of Higher Education, 29,* 1-22.

Casper, G. (1996, September 23). Criticism of College Rankings, a letter to US News and World Report. Retrieved from http://www.stanford.edu/dept/pres-provost/president/speeches/961206gcfallow.html VIEW ITEM

Clarke, M. (2002). Some guidelines for academic quality rankings. *Higher Education in Europe, 27*, 443-459.

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.

Dill, D. D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher Education, 49*, 495–533.

Dometrius, N. C., Hood, M. V., III, Shirkey, K. A., & Kidd, Q. (1998). Bugs in the NRC's doctoral program evaluation data: From mites to hissing cockroaches. *Political Science and Politics, 31*, 829–835.

Ferguson, G., & Takane, Y. (1989). *Statistical analysis for psychology and education*. New York: McGraw Hill.

Gormley, W. T. Jr., & Weimer, D. L. (1999). *Organizational report cards*. Cambridge, MA: Harvard University Press.

Huang, L., Chen, S., & Chien, C. (2015). The effect of university rankings on learning satisfaction: Social identities and self-identity as the suppressor and mediators. *Asian Journal of Social Psychology, 18,* 33-42.

Landau, S., & Leese, M. (2001). *Cluster analysis (4th ed)*. New York: Oxford University Press.

Lederman, D. (2009, June 9). The best university? Retrieved from http://www.insidehighered.com/news/2009/06/09/clemson VIEW ITEM

Lee, S. (2009, August 19). Reputation without rigor. Retrieved from http://www.insidehighered.

com/news/2009/08/19/rankings VIEW ITEM

Page, S., Cramer, K. M., & Page, L. (2008). *The sophistry of university rankings: Implications for learning and student welfare.* Paper presented at the annual meeting of the Society for Teaching and Learning in Higher Education, Windsor, ON.

Page, S., & Cramer, K. M. (2001). *Maclean's* rankings of health care indices in Canadian communities, 2000: Comparisons and statistical contrivance. *Canadian Journal of Public Health, 92,* 295-300.

Page, S., Cramer, K. M., & Page, L. (2010). Canadian university rankings: Buyer beware once again. *Interchange, 41,* 81-89.

Pike, G. R. (2004). Measuring quality: A comparison of U.S. News rankings and NSSE benchmarks. *Research in Higher Education, 45,* 193–208.

Provan, D., & Abercromby, K. (2000). *University league tables and rankings: A critical analysis.* CHEMS Paper No. 30.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom.* New York: Holt, Rinehart, & Winston.

Rosenthal, R., & Jacobson, L. (1992). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development.* New York: Irvington Publishers.

Salmi, J., & Saroyan, A. (2007). League tables as policy instruments: Uses and misuses. *Higher Education Management and Policy, 19,* 24-62.

Schmidt, S. (2004, February 2). Students 'hurt' when schools try to impress *Maclean's*: UBC accused of capping class size to improve ranking. *Ottawa Citizen*, pp. A1-A2.

Shavelson, R. J., & Huang, L. (2003). Responding responsibly to the frenzy to assess learning in higher education. *Change, 35,* 10-19.

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, Annual Meeting of the Jean Piaget Society, Toronto, Ontario.

Stevens, M. L. (2007). *Choosing a class: College admissions and the education of elites.* Cambridge: Harvard University Press.

Van Dyke, N. (2005). Twenty years of university report cards. *Higher Education in Europe, 30,* 103-125.

# Acknowledgements

# Biographies

Ken Cramer, PhD is a 3M National Teaching Fellow and full professor of psychology at Windsor, investigating educational psychology and teaching technologies.

Stewart Page, PhD (University professor, Emeritus) from the University of Windsor Psychology

Department, has investigated Maclean's rankings for over 20 years.

Vanessa Burrows is a University of Windsor psychology graduate (2015), both Sarah Mackay and Chastine Lamoureux recently graduated from Windsor with their honours degrees (with thesis), and both Rebeccal Pschibul and Victoria Pedri recently completed their 3rd year in the psychology honours program at Windsor.