The Emperor Isn't Wearing Any Clothes:

Intellectually Bankrupt Academic Prioritization

by

Professor Eric Howe

Department of Economics

University of Saskatchewan

Address: *9 Campus Drive*
                 Saskatoon, Saskatchewan
                 Canada   S7N 5A5
Phone: *306-966-5212*
Email: *eric.howe@usask.ca*

1. The Dickeson prioritization process

After a number of years being applied in the United States, the academic prioritization process of Robert Dickeson has crossed the border into Canada. Only one member of the U15—the UofS—is currently applying Dickeson's methodology, but three other universities (Regina, Brock, and Guelph) have undertaken or are currently undertaking such prioritizations.

The methodology is outlined in Robert Dickeson: *Prioritizing Academic Programs and Services*, Jossey-Bass, 2010—*Prioritizing* henceforth. We will start with a brief summary of the methodology.

The methodology involves identifying all of a university's programs and prioritizing them all. Although the methodology in *Prioritizing* is to be applied to all programs, including services, this article will only examine the prioritization of the academic programs. Moreover, this article restricts its discussion to universities although many of the following comments are equally applicable to technical schools and colleges.

On first examination, some may find the methodology to be intuitively appealing. Regardless, there is notably less appeal for many of the specific assertions in *Prioritizing*. For example, the opening sentence in *Prioritizing* is a mighty bit of mythologizing: "American higher education has been regarded universally as the best in the world" (*Prioritizing*, p. 1).

The UofS is using the methodology in *Prioritizing* as the basis for its ongoing prioritization, TransformUS, and this writer recently resigned from that process. It is important to note that the following analysis is a critique of the methodology laid out in *Prioritizing* and not the specific methodology actually being employed at the UofS. However, a few publically available UofS examples will be employed.

In these litigious times, it is noted that the following is the writer's own interpretation. Certainly, others may interpret *Prioritizing* and TransformUS differently.

Under the methodology in *Prioritizing*, the university identifies all of its academic programs, where a program is defined to be any activity which consumes resources (*Prioritizing*, p. 56). Each department would typically have many programs. For example, the Department of Computer Science at the UofS has twelve: eight undergraduate, three graduate, plus a separate program for research.

Each program is to be evaluated according to criteria. *Prioritizing* recommends the following ten criteria (*Prioritizing*, p. 66):

> 1. History, development, and expectations of the program;
>
> 2. External demand for the program;
>
> 3. Internal demand for the program;
>
> 4. Quality of program inputs and processes;
>
> 5. Quality of program outcomes;
>
> 6. Size, scope, and productivity of the program;
>
> 7. Revenue and other resources generated by the program;

8. Costs and other expenses associated with the program;

9. Impact, justification, and overall essentiality of the program; and

10. Opportunity analysis of the program.

Each criterion is assigned a weight reflecting its relative importance, depending on the university (*Prioritizing*, p. 69). At the UofS, the smallest weight is for Criterion 1 (5%) and the largest is for Criterion 5 (18%). Information, both quantitative and qualitative, is collected about each program relative to each of the criteria, some provided centrally and some by program proponents.

There is some discussion in *Prioritizing* of how to score the individual criteria based on the information collected (*Prioritizing*, p. 95-99); then the overall score of the program is arrived at as a weighted sum of the results for the individual criteria scores for each criterion. That provides a prioritized list of academic programs from the "stars" down to the "dogs" (*Prioritizing*, p. 89).

Then the programs are divided into categories based on the percentile of their aggregate score. *Prioritizing* discusses quintiles (*Prioritizing*, p. 102), though it also discusses other divisions such as quartiles. Programs in the bottom category are candidates for elimination and those in the top, for enrichment. Different labels are discussed for the intermediate categories. The UofS has adopted quintiles with the following interpretations:

Quintile	Programs per Quintile	Interpretation
1	Highest 20%	Maintain the program with additional resources
2	Next to highest 20%	Maintain the program with existing resources
3	Middle 20%	Maintain the program with reduced resources
4	Next to lowest 20%	Candidate for restructuring or reconfiguration
5	Lowest 20%	Candidate for elimination

The reader should note that the bottom quintile makes up one in five of the programs, so the UofS has taken the position at the outset—prior to gathering data about the programs, prior to scoring, prior even to deciding on priorities and weights—that one in five of the programs will be candidates for elimination. Moreover, three out of five programs are assumed to not be worth what they cost.

The following critique will point to a variety of logical flaws in the methodology in *Prioritizing*. Any one of those flaws alone is sufficient to vitiate the methodology. Taken together, they show that the methodology is intellectually bankrupt.

2. The size bias

Prioritizing scores programs using the weighted sum of priorities, as discussed in the previous section.

Samuel Langhorne Clemens (Mark Twain) famously advised writers to "use the right word, not its second cousin." His point was that there is a vast difference between using the right word and using a

word which is almost right. Similarly, there is a vast difference between prioritizing with the right scoring function and using a scoring function which is almost right.

The correct objective for academic budgeting is obvious. The university should maximize the benefit (broadly defined) from spending its budget so it needs to allocate spending according to what yields the maximum benefit per dollar of cost. It is tautological that program scoring should be done on the same basis: as a ratio of benefit to cost.

By using the weighted sum of priorities, *Prioritizing* scores using a second cousin. The two approaches are inherently different because a ratio of benefit to cost is different than a weighted sum.

The effect of the difference introduces biases which strike at the heart of what a university is. One notable bias is one in favour of large programs. To see how small programs are treated unfairly, return to the list of criteria in the previous section. All but Criteria 1 (History) and 10 (Opportunity) involve program size either explicitly or implicitly. Four of the criteria explicitly involve size: Criteria 2 (External demand), 4 (Internal demand), 6 (Size), and 9 (Impact). Two others, Criteria 7 (Revenue) and 8 (Costs) also explicitly involve size but will be dealt with separately.

Consider what happens when a program is hypothetically doubled in size. Start with a program for which Criteria 7 (Revenue) and 8 (Cost) cancel each other out relative to the assigned weights. Double everything—the number of teaching awards, the number of faculty, the number of students, the amount of tri-council funding, the number of publications, and so forth. The benefit per dollar of cost is unchanged by the doubling, so the logically correct evaluation of the program would be unchanged. But the weighted sum of criteria would move the program to a higher score. Similarly, multiplicative reduction would move it to a lower score. All for the same program with an unchanged benefit per dollar of cost.

The reason for selecting a program for which Criteria 7 (Revenue) and 8 (Cost) cancel out is straightforward. If a program is a cash cow then doubling would provide additional resources for other programs, so it could be appropriate for it to move up in the rankings. If the program is a resource hog then doubling could move it down and that too could be appropriate.

Now consider programs for which Criteria 7 (Revenue) and 8 (Cost) do not cancel each other out. For such programs, large programs still have an advantage, though there is a trade-off between their size on one hand and their cost versus revenue on the other.

Certainly some will state that those doing the scoring could somehow take the size bias into account and adjust their scores appropriately. Whether they can and will do so while scoring hundreds of programs seems highly questionable, but perhaps this becomes an empirical question.

Empirically, how large is the size bias? At this point in time the results for the UofS have not been publically released so it would be inappropriate to reveal whether the rankings are biased with large programs doing better than smaller. At the time that they are released, it will be appropriate to return to this question to illustrate the point.

At that time, some will doubtless state that larger programs simply turn out to be better than smaller ones—perhaps because they have been more successful in obtaining funding in the past. But note that

the large number of programs will allow for fine empirical distinctions. In the empirical analysis, the ranking of smaller programs can be compared to larger within a given department or college, somewhat correcting for quality.

The size bias is exaggerated by the focus on programs rather than departments because the variation in the size of programs is greater. Each department is made up of a variety of programs so summing over them for the department reduces the proportionate variability in size.

The size bias strikes at the heart of what a university should be. Universities are meritocracies so more advanced programs tend to be smaller. This is particularly easily observed at the department level: the graduate programs will be smaller than undergraduate programs with the PhD program smaller than the Master's. There are usually only a handful of post-doctoral students. A bias favouring large programs is a bias against merit.

In terms of departmental development, a department facing *Prioritizing* would be well-advised to strive to be large and mediocre rather than small and excellent. (Of course, large and excellent would be better yet, but that may not fit the budget.)

Universities are vectors of social change—that is one of their vital roles. Another consequence of the weighted criteria scoring function is to bias the prioritization against programs which are involved in the process of social change. Before developing that point, we turn to showing that academic programs are not the appropriate basis for academic prioritization.

3. Academic programs should not be used to prioritize a university

The *Prioritizing* methodology uses programs—rather than departments—as the basis for prioritizing. However, the complicated relationships between the programs within departments make it impractical to prioritize the programs for a university. For universities, the use of programs, by itself, invalidates the methodology.

The relationships between programs within departments are rich and varied across a university. As an example, consider the relationship between the PhD and the research programs. For some departments the two are hand in glove. But there are other disciplines where the two programs are distinct.

One size doesn't fit all. At one Town Hall on prioritization at the UofS, a professor from the Department of Computer Science pointed out that it was conceptually impossible to separate the research and PhD programs in computer science and asked why prioritization was doing so. The answer was that in other departments it was conceptually impossible to combine them.

Similar problems apply to some departments' Master's and PhD programs. Some departments, for discipline-specific reasons, award Master's degrees only to students unable to complete the PhD, so there is really no point in evaluating the Master's degree at all. If the PhD is to be expanded, contracted, or eliminated then the same fate should befall the Master's. On the other hand, other departments treat the Master's as a distinct entity so the Master's and the PhD can perhaps be evaluated separately. Even then, some courses will typically be shared, making the measurement of cost difficult to impossible.

It would be conceptually possible to deal with this difficulty by treating programs differently, combining say the research and PhD programs for some departments and not for others. However, it would be difficult to do for the hundreds of programs within a university.

Measuring and using the cost of a university program, on the other hand, is confounding even though Dickeson blithely asserts that "Costs are readily measureable" (*Prioritizing*, p. 65). For example consider the cost of delivering undergraduate introductory courses which are taken by people who will ultimately graduate from any of a variety of different programs. For those who will ultimately graduate within the department, expenditure on delivering the course has to be proportioned among a number of programs within the department. It is necessary to do the proportioning without waiting for students to graduate since otherwise cost data would not be timely—delayed by years of waiting for graduation.

Imagine the data problem faced in proportioning a department's expenditure among programs. Most faculty would find it extremely challenging to decide personally how much time they put into each of the department's programs: how much time for research as opposed to teaching students who will ultimately graduate in the Honours Bachelor's program for example. And how do you count those periods with late nights leading up to a deadline? To decide these issues for each of the faculty members in an entire department would be daunting.

As noted above, the Department of Computer Science at the UofS has twelve programs. Each, on average, uses 8% of the department's expenditures. But what proportion would go specifically to, say, the 4 year Bachelor's in bioinformatics? What about Honours?

There are two immediate consequences of this data estimation conundrum. Some individuals will simply report their best guess, which is sometimes inaccurately referred to as a best estimate. Others will attempt to game the system by strategically reporting what they believe will help their cause: a program which is possibly on the chopping block may be reported as having minimal cost, its share of expenditure having been moved into another departmental program which is thought to not be vulnerable. (Strategically, the underreporting of cost for a vulnerable program is extremely attractive both because it makes the program appear to be more cost effective and also because there will be less of a hit to the departmental budget if the program is ultimately eliminated.)

This nature of the program data makes accurate prioritization of programs impossible. Although the prioritization process is to identify outstanding programs, no program will likely be outstanding according to all ten of the criteria. Rather an outstanding program will be extraordinary in some ways but not others. For sake of argument, suppose that "extraordinary" according to a particular criterion means that the program is in the 99th percentile relative to that criterion. In the process of prioritizing, when a program is encountered which appears to be in the 99th percentile, the above considerations suggest that the program is probably not unusually good, but rather the data are unusually bad. Which is more likely, that the program is actually extraordinary which happens with 1% probability or that the data are in error which happens frequently? Most will conclude that it is the data.

4. Just because a program is in percentile x doesn't mean it should be a candidate for elimination or enrichment

The treatment of quintiles adds to the panoply of illogic in *Prioritizing*. After the programs are scored, they are divided into groups by their percentile score. As discussed in Section 1, different percentile groupings are discussed—quartiles for example. However, quintiles feature the most prominently in *Prioritizing* so this discussion will use them. Programs in the bottom quintile are candidates for deletion whereas programs in the top quintile are candidates for enrichment (*Prioritizing*, p. 96).

However, there is no reason to suppose that a program in the top quintile should *a priori* be a candidate for increased funding. Admittedly, one criterion used in the methodology is opportunity analysis which delves into the how the program would benefit from increased funding (*Prioritizing*, pp. 86-87) but it is only one part of one of ten criteria used in the linear sum for scoring. Just because a program is a "star" does not mean that it should be a candidate for increased funding because there is no reason to suppose that it has an important use for further funding. In fact, it may have benefited from previous over-funding and be a candidate for reduced funding.

Similarly, just because a program is a "dog" doesn't mean *a priori* that it should be a candidate for elimination. A university has many priorities and a program which is weak can rationally be kept because of those priorities. In fact, depending on the university's priorities, a program in the bottom quintile may be best treated as a candidate for enrichment so it can be strengthened.

(The reader is reminded that the characterization of programs as stars and dogs is from *Prioritizing*.)

Universities are meritocracies, so I suspect that some readers may still believe that programs in the bottom quintile should be candidates for elimination. Faculty (especially those who curve grades) may see a metaphor in their own grading with weak students eliminated. Why shouldn't a weak program automatically be a candidate for elimination? As noted in Section 2, one of the central roles of universities is to serve as vectors of social change. It is straightforward to imagine a program which is weak academically and that academic weakness puts the program in the bottom quintile even though the program is extraordinarily important for social change (or indeed for any other university priority).

Consider Saskatchewan. There is a gap between the levels of educational attainment of the province's Aboriginal and Nonaboriginal populations. The gap is growing despite the increasing level of educational attainment of the Aboriginal population because the educational level of the Nonaboriginal population is increasing even faster. The Aboriginal education gap is one of the contributing factors to high levels of unemployment, welfare dependence, crime, FASD, HIV, and other social problems which are present in the province's Aboriginal population. Moreover, Saskatchewan will be majority Aboriginal in the foreseeable future so the province has to attach an extraordinarily high priority to Aboriginal education, and the UofS should attach a similarly priority. The University agrees, as is evidenced by the very prominent positioning of Aboriginal initiatives in its planning documents. But a program representing an Aboriginal initiative can easily fall in the bottom quintile and—following the methodology in *Prioritizing*—become a candidate for elimination.

5. A budget process without a reasonable measurement of cost is silly nonsense

Proportioned expenditure is unrelated to cost. Consider for example the Post Degree Specialization Certificate, a small program some departments offer to allow alumni to return to university and improve their educational credentials. In considering eliminating the program, the cost of the Post Degree Specialization Certificate is not the proportion of expenditure on students in the program. Instead, it is much smaller if the only consequence of its elimination would be a few empty seats in a handful of classes that are offered anyway to those in other programs. Programs are defined in *Prioritizing* to be any activity which consumes resources (*Prioritizing*, p. 56), so any program will by definition have a proportionate share of expenditure but may have minimal cost.

Again, consider the Department of Computer Science where, as is typical, the course requirements for the programs consist of intersecting subsets. For example, the only additional course requirement for the honours in bioinformatics beyond that for the four year major in bioinformatics is one 3-credit course. Elimination of the honours program would not save the proportional share of expenditure on students who will ultimately get honours, but would be largely limited to the cost of that one course.

Prioritization is a budgeting process! It is hard to sufficiently emphasize the obvious: a budgeting process without a reasonable measurement of cost is ***humiliatingly nonsensical!***

6. Other comments

There can be little doubt that the methodology in *Prioritizing* can be extremely attractive to some in senior administration. By making 20% of programs candidates for elimination, a great deal of power is created and then bestowed on those who decide ultimately whether the trigger should be pulled. Personal testimonials by administrators, of the sort displayed prominently in *Prioritizing*, should be taken as special pleading by individuals with an axe to grind.

It is impossible for this writer to read *Prioritizing* without being struck how the methodology is informed by a community college view of the world. If it were necessary to decide on the priority for—say—the Montgomery Community College of their programs in driver training versus catering then the methodology in *Prioritizing* might work acceptably well. Not so much for a research university's complicated and interdependent network of programs.

Each of Sections 2 through 5 present independent arguments that each vitiate the methodology in *Prioritization*. Any one of the arguments is sufficient. Taken together, they show that it is intellectually bankrupt.