The Academic Affairs Point System for Evaluating Graduate Programs

Scott Farrand

The Academic Affairs point system, which evaluates graduate programs for potential enrollment reductions, assigns points in three general categories: Program Quality, Program Need, and Program Cost/Benefit. This analysis follows that organization, and concludes with an analysis of some overriding methodological issues.

Measurement of Program Quality

Apparently, Program Quality was initially planned to be measured along 8 dimensions. For some of the senate's recommended criteria, data were more difficult to obtain, and these criteria were not used, including Quality of the Faculty (degrees held, scholarly activity, and teaching effectiveness), Alumni Success, and Special Action Admit Rate. Also, two of the criteria (Course Offerings and Faculty Involvement) were determined to have been met by all programs, and so they did not play a role in separating programs. There remained then three criteria by which program quality was measured: Course/Program Development, Admissions Requirements, and Admissions Decisions.

Course Program Development was measured by whether Form A or Form B has been submitted over the past 3 years, and the range of scores in this category was 0-1.

Admissions Requirements were measured according to the gpa requirement of the program (1 point for a gpa requirement of at least 3.0, one point for requiring a test such as the GRE of applicants, and one point for other requirements such as letter of recommendation or writing samples), resulting in a score of 1-4 in this category. A requirement of the equivalent of a bachelor's degree in the discipline did not yield any points.

Admissions Decisions were measured simply by the acceptance rate (ratio of applicants recommended for admission to total applicants), with a ratio of less than 35% yielding 5 points on this 1-5 scale. Applicants that did not meet the admissions criteria for the program were included in the denominator of this ratio.

The points allotted to the measurement of program quality were doubled, yielding an effective combined range of 4-18 for the categories of Admissions Requirements and Admissions Decisions. As a result of this weighting, 42% of the total evaluation of a program was based on these two categories.

Analysis

A major weakness of the system in measuring Program Quality is probably the inappropriateness of its application across disciplines. Programs that, because the nature of the discipline. require the equivalent of a bachelor's degree (because the graduate courses are unapproachable without the very specific undergraduate preparation) are less likely to need supplemental admissions requirements. Thus they are not only punished by not being granted any points in this category for their specific undergraduate coursework requirements, they are further punished for not

having supplemental admissions requirements that might simply not be germane to their admissions decisions. The difference between a high score and a low score in this category (Admissions Requirements), based entirely on the nature of the discipline, could more than bridge a program from one recommended for a suspension of fall admissions to a program recommended for no reductions in fall admissions.

Programs that provide greater clarity in their admissions requirements are at a disadvantage in the Admissions Decisions category. The more clear-cut the admissions requirements are, the less likely it is that applications will be received that do not meet the admissions requirements. Without a closer examination, it is difficult to determine how this consideration might affect the points awarded in this category. This is another example of a discipline-based variable however, as graduate programs that are by their nature more interdisciplinary might be more likely to receive applications that would require more nuanced consideration of backgrounds, while graduate programs that are more of a direct continuation of specific undergraduate education could easily attract a smaller pool of applicants because the applicants have a more clear means of self-evaluation for admission.

In some disciplines, these two categories (Admissions Requirements and Admissions Decisions) might be very useful measurements of program quality, and would be useful in making distinctions among such programs, but these are not very useful in comparisons across disciplines, given the differences that are likely to exist solely because of the discipline. If this data were collected at several CSU campuses, and an analysis of variance were performed, we might learn more about what is actually being measured. There certainly is reason to suspect that what the points are revealing about the programs is really more about the nature of the disciplines than the quality of the programs, and would yield similar separation of programs at comparable CSU campuses.

Given that the weighting of these two categories overwhelms the weighting given to any other categories (providing more than double the points needed to bridge between suspension of admissions to no change to admissions), the implementation of the senate criteria in these categories brings the instrument into serious question.

Measurement of Program Need

The Faculty Senate provided examples of several measurements that might be considered in evaluating Program Need, including the number of applicants, number of admits, number enrolled, and employer surveys. Just one category was used in the Academic Affairs point system for evaluating Program Need, the ratio between enrollments and admits, which was turned into a scaled score, 1-5.

Analysis

While this ratio does seem relevant to program need, it only measures the need for the program through the actions of students. A major need for a program in the community is not necessarily manifested in a major demand for it among students. A program could possibly be crucial to the University or to the community without causing students who are admitted to accept at a high rate (if for example universities in other communities are competing with Sac State for these graduate students).

Measurement of Program Cost/Benefit
Measurement of these criteria was proposed in six categories, and was actually performed in five categories (Grad FTES/Total WTU; Total FTES/WTU; Time in Program; TA/GA Positions; Interdisciplinary), with one category not being evaluated (Replicative Components, a proposed review of the catalog for repetition of courses or other components of programs).

Graduate FTES/WTU is a measurement of the cost per student for providing graduate coursework in the program (reflecting the size of graduate classes). Total FTES/ WTU is a measurement of the cost per student for providing any coursework, graduate or undergraduate, in the program. These two categories were each scored on 1-5 point scales. Time in Program reflects the average time to complete the program, and was scored 1-4. TA/GA Positions relates to whether the program offered assistantships that provide skilled labor to the university that itself might provide for decreased costs. This was valued at 0-1 points. Interdisciplinary programs are considered for the possibility that they provide efficiencies as they serve more than one program, and were given 0-1 points.

Analysis
The vast majority of the weighting for Cost/Benefit is given to categories that primarily measure class sizes (Grad FTES/Total WTU and Total FTES/WTU). This choice makes it inherently more important that program quality need be properly measured. Small classes in a program might in some cases be an indication of a program in trouble, while in another program it might be the nature of the discipline, if it requires more individualized instruction. The question of whether programs that require more individualized instruction should reduce admissions must depend on the Quality and Need for the programs, and so systematic weaknesses of the Academic Affairs point system in those areas become all the more significant.

A simple calculation of these ratios does not come close to producing an analysis of the cost/benefit ratio for enrollment in the program. The real question is how much money would be saved or lost if enrollment were reduced. If reducing enrollments produces a reduction in WTUs required by the program, then that is a very different situation from one in which reducing enrollments just produces smaller class sizes with the same number of sections. Certainly WTU/FTES ratios tell us something about the cost of the program per student, but it is another thing altogether to translate that into a statement about the savings that might be produced by reducing enrollments, short of program termination.

In areas of Program Quality and Program Need, there is ample room for disagreement about the relative weighting to be given to different categories of evaluation. In Cost/Benefit analysis it is self-evident that the weighting of the categories ought to relate to some actual measurement of cost benefit, so that if a category might relate to savings of a small amount, it would be weighted less than a category that measures greater savings. Although the FTES/WTU calculations (2-10 points) do relate to sizeable expenses, the TA/GA consideration (0-1 point) in some departments relates to much greater costs. Cutting enrollments in some programs would produce little to no savings in expenses related to teaching the graduate classes, but would incur major costs when the teaching assistant or graduate assistants who perform crucial tasks are replaced by more expensive alternatives such as part-time faculty. Impact on external funding is another major

consideration associated with reductions in graduate students who might work on grant-related projects.

Weighting the TA/GA category with 0-1 point while giving the FTES/WTU categories both 1-5 points might be entirely upside down, in terms of relationship with Cost/Benefit, which is what these were intended to measure. Absent an analysis of the impact on reductions in enrollment on the actual costs to the University in each of these categories, we cannot know how these categories relate to Cost/Benefit, but the weightings appear arbitrary and disconnected with their purpose.

Overall Analysis

The assignment of weight to categories appears inconsistent with any overall consideration of the relevance of the categories to the overall evaluation of programs. Perhaps there is an explanation for the weightings, but Academic Affairs has offered none, other than the doubling of points in the area of Program Quality. But how were the weightings assigned to the categories before the points were doubled? Need was given 1-5 points, while Cost/Benefit was given 3-16 points, and after doubling Program Quality was given 4-24 points (and since all programs received the same points in two of those categories, the effective range was 4-20 points).

As the process was described to GSPC, the range of points given in a category sometimes was determined by the range in data rather than the importance of the category to the overall evaluation. If it was an all or nothing category, it was worth 0-1 points. If the data in a category appeared in several clusters (e.g., categories in which the data was some sort of ratio), then the clusters were made to correspond to points, and the category might get 1-5 points. This is a serious and ironic methodological error, hardly the sort of thing that should be abided in a determination of which graduate programs might be suspended.

There are aspects of the point system that might be useful in comparing a program at one institution with another in the same discipline at another institution, but there is little to support their use in comparisons across disciplines, much less as a basis for major decisions about future directions for the campus.

In describing this exercise, Provost Sheley repeatedly commented that "interesting patterns emerged," although we have not been told what patterns emerged. The presence of interesting patterns immediately brings to mind the question of whether those patterns are fact or artifact. Are they a result of the population or of the instrument? Do those patterns tell us more about the graduate programs that were evaluated or about the point system that was used to evaluate them? It appears that the Provost did not give sufficient consideration to the possibility that the instrument was seriously flawed before he presented the deans with proposals calling for suspensions of admissions resulting from the instrument.

None of this is to say that the point system could not be improved to remedy the flaws outlined above. It is also clear that this analysis is limited by the perspective I have, coming from my discipline. Anyone designing a point system such as this would be prone to make some systematic errors for lack of perspective. The development of a valid point system begs for input from a broad spectrum of academic disciplines.