

Freeman Committee Interim Report Overview

Premises upon which the Committee conducts its work:

- 1) Graduate education and research cannot be separated
- 2) To achieve our research aspirations graduate programs must be strengthened.
- 3) Our current budget model promotes quantity, rather than quality of graduate education
- 4) Programs vary dramatically in quality and so differential allocation of resources should promote quality
- 5) Focus is development of processes for differential resource allocations.

The focus of this committee's reach is restricted to doctoral education. Many of these same issues impact masters' and professional education. However, these programs exceed the charge of the committee.

Budget Issues driving this analysis:

- Many graduate programs make extensive use of graduate assistantship appointments as GTAs and GAAs; and less use of GRAs
- GA appointments require colleges to provide fee authorizations with initial distribution of resources (at rebasing) based upon historical use regardless of quality or appointment type
- Many programs rely on GTAs to attain their undergraduate instructional goals
- In research doctoral programs (not professional programs), the universal gold standard across the country is that the students are not expected to pay the cost of their education themselves, but, rather, this cost is usually borne by the institution, research grants, etc. Professional graduate programs can and do charge tuition at varying rates. However, tuitions for PhD programs are determined at the state level.
- Funds for fee authorizations were distributed to colleges during budget restructuring according to historical usage. Any increases to college budgets since then have not included increases for this purpose, but have been based on other criteria (e.g., undergraduate enrollments). Centrally supported fee authorizations in college budgets have not been incremented from their 2002 tuition base
- Graduate tuition has gone up 35% since 2001
- There are system wide caps in Ohio on doctoral subsidy that do not permit the institution to receive additional subsidy if the numbers of doctoral student FTE go beyond a specific level. This means that to add new programs or increase the size of some, there might need to be a corresponding decrease in others in order to avoid a drop in revenue from the state.
- State Share of Instruction (SSI) has been flat or declining (and has been taxed) while tuition has gone up
- Units are not receiving revenues sufficient to pay for graduate programs
- In some disciplines external support for GAs is non-existent

- In other disciplines, the cost of a GA (stipend, benefits, tuition) approaches \$30,000 and for many faculty hiring a postdoctoral fellow has become an attractive option. Similarly, some departments are now finding it more financially attractive to appoint lecturers rather than Graduate Teaching Associates.

The special problem of interdisciplinary programs

- For Interdisciplinary Graduate Programs, revenues accrue to participating colleges but costs are born by programs
- There is a fundamental disconnect between interdisciplinary graduate education and interdisciplinary research/centers/institutes

The Committee proposes a variety of funding strategies to promote outstanding doctoral education, including program review and differential funding based on quality, as well as initiatives intended to create a funding model that fosters quality in doctoral education

These include:

- 1) identify a set of metrics to assess quality of PhD programs
 - a. proposed metrics outlined in the interim report were discussed at Council of Deans on 11/18/2004; these will be used to propose a model of resource allocation based upon quality
 - b. data collection for several PhD programs is currently underway
 - c. we will assess the model output against our perceptions and against results of recent external reviews/accreditation exercises to judge its validity
 - d. assuming validity, we will propose how the model might be used more broadly
- 2) Study and propose new budget models for funding doctoral education; key elements to be studied will include
 - a. differential tuition by student status (e.g. pre-and post candidacy)
 - b. differential allocation of fee authorization across programs
 - c. differential allocation of fee authorizations for GRAs versus GTAs and GAAs
 - d. differential tax on graduate subsidy and tuition based upon program quality
 - e. A Selective Investment program for graduate education – taxing some programs and investing it in others
 - f. Costs must be identified and a strategy to pay for them developed
 - g. The new budget model should be explicitly connected to graduate education quality and the research metrics that drive this quality
 - h. Incentive programs that recognize and support entrepreneurial activity (e.g. fee authorization support for programs with substantial external support)

Interim Report

The Committee has met 5 times, starting in August of 2004, with additional individual work pieces from each Committee member. This *interim report* does not attempt to deal with many of the larger issues implicit in the Provost's Charges, such as a comprehensive plan for the funding of graduate education. Rather, to this point the Committee has concentrated on laying the ground work for implementing a process for judging the overall quality of OSU's PhD programs, with the implied recognition that there is no strategy available for continuing all of OSU's PhD programs at their current levels *while simultaneously assuring* that a substantial portion of the programs gain the level of national prominence envisioned in the Academic Plan.

The Committee first analyzed the Charges in detail, and readily agreed that Charges 1-4, which fundamentally address how to realistically determine the quality of a given graduate program, can only be systematically approached by the establishment of a group of metrics* focused upon the quality of the student within a program, and the student experience within the program through the quality and engagement of the faculty.

Of paramount concern was the recognition that of the some 100 Ph.D. programs within the University, there is remarkable diversity in terms of quality indicators of excellence. Indeed, the Committee has spent the majority of its time identifying the largest set of metrics that could be applied University-wide, while also specifically calling out procedures for ranking those programs which have no obvious campus comparative (e.g., comparison with CIC equivalents when necessary†).

The substance of this interim report is an outline of the metrics for judging the quality of PhD programs at OSU, with extended notes indicating the limitations and potential

* The Committee devoted considerable discussion to what constituted a set of measurable outcome metrics, as opposed to those which addressed process only. The Core Metrics presented below were chosen with the aim of measurable outcomes attached to each.

† The Committee has realized that many metrics for more specialized programs will have to be cross-institutional in nature. This will involve considerably more effort on the part of the group that is tasked with carrying out these recommendations, and some non-trivial costs.

pitfalls in employing these metrics without considerable care. The Committee has not yet taken up the details of implementation for gathering the data in a timely manner, nor the process of the basic formula by which these data will be used in model that would yield a sensible and reasonably accurate assessment of OSU's PhD programs[‡].

The Committee has noted that given the time and monetary restraints imposed upon this proposed analysis, the best result that can be expected from application of these metrics is likely to be a sensible grouping of programs into bands[§]. The top bands would presumably include programs that should be encouraged to continue in their drive for national and international recognition. The middle bands would presumably be those that are judged as either too new to OSU to rate, are of such value to OSU's mission that they must be supported, or are undergoing obvious improvement and should be encouraged to examine their programs with care in order to emulate the successes of the top tier programs^{**}. The bottom bands would presumably be those which unable to make a convincing case that they are (a) of special value to OSU's educational mission, (b) have not been historically marginal with little or no improvement, or (c) essential to OSU's future in terms of the Academic Plan.

It is not within the Charge of this Committee to go beyond the Provost's instructions: The Committee fully recognizes the difficulty in ultimately assigning programs into those that will be supported, and those that will suffer cutbacks. Our purpose is to provide a mechanism or a tool for the Provost or her designees to undertake an extraordinarily difficult task. Yet the Committee is convinced that some process, whether or not the one

[‡] However, as indicated in the section below describing "next steps", the Committee proposes to gather the required information on a handful of graduate programs to assess the feasibility of applying the proposed metrics.

[§] Any attempt to actually rank the 100 PhD programs in quality (that is, 1->100) is subject to a level of scrutiny and debate that the Committee views as unnecessary at best, and probably futile at worst.

^{**} The Committee has discussed the problem of acquiring metric data over some extended period of time in order to judge the "trajectory" of a program. This makes the gathering of data more difficult compared to "snap shot" analysis of programs for a given year. There appears to be no alternative to analyzing program data over a period of time on the order of 5 years.

proposed here, is essential for OSU to move a subset of its graduate programs into the very top echelons of national research universities^{††}.

The Committee first discussed at length whether there were processes already in place that would satisfy the Charges, or examples of review processes in other institutions that could be easily ported to OSU. The Committee adopted the position that a satisfactory response to the Committee Charges would have to involve a campus-wide process that was transparently constructed, widely reviewed, judged by the faculty as being as fair as possible, yet capable of being implemented in a relative short period of time (less than a year^{‡‡}):

1. There appear to be no satisfactory review mechanisms for graduate programs within OSU that would satisfy the charges. (Senate Analysis of Graduate Programs, Senate Report on Barriers to Interdisciplinary Research) The Committee considered plans announced by Vice Provost Randy Smith to start systematic reviews of graduate programs at a rate of some 5-10/year and concluded that a much faster process was needed to meet the requirements of the Provost.
2. The generally accepted method of graduate program review is to ask for external committees to visit the program on campus, and to write a detailed report which addresses strengths and weaknesses. While this process is followed by several OSU colleges in analyses of their individual graduate programs, and is dictated by accreditation bodies in others, it is not systematically applied across the campus in a manner that is applicable to the Committee's Charge. (One of the more effective of such procedures is the in-depth rolling review every 7 years of all programs at

^{††} Essentially, OSU must undertake some process of self examination of its PhD programs that produces a reasonable approximation to a quadrant graph of quality vs. importance for its graduate programs. Insufficient resources exist to maintain our current number of programs and to make improvements called out in the academic plan.

^{‡‡} To this point the committee has not solicited input from any stake-holders on campus. In the section on "next steps" the Committee suggests a wide dissemination of its current thinking, with a specific goal of receiving necessary and useful feedback on its directions.

Northwestern University. This process, while capable of fully responding to the Committee's Charges if implemented across the University as a whole-and is the model for Vice Provost Smith's process, would take considerable time and effort to implement, as well as substantial funding.)

3. Proposed Data for Constructing a Model for Analysis of PhD Programs at OSU:
The Committee has assembled and discussed at length 7 categories of Data to be gathered from the programs in order to construct a reasonable model to analyze the PhD programs in a time period of 3-4 months. We have further added 3 Supplemental metrics which we recognized were clearly not applicable widely across all programs, but could be valuable metrics when compared to the University's aspirational peers. Care should be taken to note that even within the "core" 7 metrics, allowances for differences in the importance of the several of the metrics by discipline are paramount. Again, the most meaningful comparison will be with the equivalent programs in the University's aspirational peers.

The data to be gathered from each program are, to the extent possible, consistent with the NRC data that will be required for the upcoming National Rankings. The committee proposes that data be obtained on standard forms. Further, the Committee proposes the creation of an OSU internal-access-only web site where all of the general data for each program will be posted for examination by each program faculty in order to assure accuracy, and to have various levels of depth within the web site for each program.

Next Steps:

The Committee acknowledges that while compiling a list of Metrics to use in measuring the strength of a program is relatively straight forward, actually implementing the process of gathering the data in a reasonable time and with affordable effort may well be another. The Committee proposes, between the submission of this interim report and the due date of the final report, to try to gather data on a small subset of programs to test the feasibility of the process. We will then construct a model that assigns programs to the three bands

and assess the model's success according to our *a priori* assessments of program quality.
§§ This exercise will help us identify which metrics are either redundant or of little actual use in discrimination of programs.

The Committee, upon acceptance of the interim report by the Provost, proposes to initiate a series of discussions with various faculty and student stake-holders around the campus. We propose these groups be identified in collaboration with the Provost. These discussions would be designed to gain acceptance of the process as being fair and even handed, and to obtain feed-back and suggestions on the process.

As discussed above, the Committee must also address some of the larger issues called out by the Provost, notably the issues of aligning the costs of graduate education with the resources, providing guidance on whether the graduate program should become larger or smaller, whether tuition stipends should be re-centralized for competitive bid by the programs, and finally whether the campus should adopt a policy of writing tuition into grants whenever possible.

7 Core Metrics

- I. Judging the quality of entering graduate student within program¹:**
- a. Primary indicator is GRE scores (both general and subject specific, if offered)
 - b. Quality of UG institution (as roughly determined by USNEWS)
 - c. Undergraduate GPA (possibly normalized by approximate-within 25%-USNEWS ranking of undergraduate school)
 - d. Ratio of national to international students admitted and/or enrolled compared with similar programs at our aspirational peers
 - e. A combination of:
 - i. What is ratio of applicants to total graduate student number (high is better)
 - ii. What is the ratio of admits/applicants for the program (low is better)

^{§§} The Committee proposes to choose two programs from each of the committee members Colleges and to use as much data as possible already within the data storehouse as overseen by Julie Carpenter-Hubin.

- iii. What is the ratio of enrolled/admits for the program (high is better)

II. Time to Degree and Graduation %^{II}

- a. Vary, dependent upon program; care to compare OSU units to University Peer Aspirational Institutions
- b. Distribution (median vs. mean and higher moments) more meaningful than one number
- c. Master required/yes/no (separate programs in the analysis?)

III. Systematic Application of Standard Graduate Reports^{III}

- a. Comparison of results across all programs to University averages on Graduate School Exams as compiled by Graduate School

IV. Percent of students within a given program receiving a Fellowship^{IV}

- a. Only Fellowships to count are competitive, non-departmental, non-College.
- b. Examples:
 - i. University-wide as administered by Graduate School
 - ii. National Fellowships (e.g., NSF, NIH, Sloan, Fulbright, etc.)

V. Training Grants within Program^V

- a. Applicable only to programs eligible (compare University Peer Aspirational Institutions)
- b. Historical as well as current success in obtaining Training Grants

VI. Ratio of GTA/GRA within program^{VI}

- a. Highly dependent upon program; meaningful comparison only by using University Peer Aspirational Institutions Data
- b. % of GRA's tuition supported by non-University sources
 - i. Program specific, compare University Peer Aspirational Institutions
 - ii. % of students where stipend is obtained externally and tuition is on supported by OSU tuition authorization

VII. Faculty Quality Indicators^{VII}

- a. Use of NRC Gini Coefficients to measure:
 - i. Publications per graduate faculty
 - a. Quality of journals
 - ii. Citations per graduate faculty
 - iii. Extramural support per graduate faculty
 - iv. Graduate Student/faculty ratio

- 1) Distribution of faculty who actively supervise graduate students
- b. % of Faculty who are externally recognized outside of department
 - i. External Recognition
 - 1) Fellows of Professional Societies
 - 2) Major award winners (e.g. Sloan Foundation Scholars)
 - 3) Appointments to National Level Boards
 - ii. University Recognition
 - 1) University wide honorifics:
 - a) Distinguished Scholar
 - b) Distinguished Teacher
 - c) Distinguished University Professor
 - 2) College-wide honorifics
 - College Distinguished Professor
- c. Number of Associate Professors and years in rank

SUPPLEMENTAL METRICS (Heavily Program Dependent)

I. Student Professional Activity while in Program

- a. Program specific, e.g.:
 - i. Presentations at professional meetings
 - ii. Performances
 - iii. Papers published
 - iv. Grant applications written
 - v. Grants received

II. Where do the Graduates go after completion of degree:

- a. Initial position (program specific)
- b. After 5 years
- c. Comparison program by program to University Peer Aspirational Institutions

III. Uniqueness of Program

- a. How many similar programs exist:
 - i. In University Peer Aspirational Institutions
 - ii. In the World
- b. For small programs:
 - i. Balance between quality uniqueness and simultaneously being in the top 5 and bottom 5 programs in the world

NOTES ON METRICS:

I Quality of Student admitted.

a. GRE. This is the most objective normalized metric to compare quality across many confounding variables. However, it is only one aspect of preparedness and potential of candidates, and must be viewed in the context of other objective and subjective measures. Disparities can exist with UG experience and performance tied to poor standardized testing skills. Nevertheless, a minimal threshold should be identified as desired of students in each program, such that exceptions are examined closely to ensure success.

b. Quality of UG institution. This is a good objective and subjective measure that must be used in combination with the specific major and program of training, which can vary in strength at each school, as well as GRE and GPA.

c. GPA. This is an objective measure that is strongly confounded by the institution and course of study. However, it can indicate strengths underrepresented by GRE. As with GRE, a minimal threshold should be established for each category of school and courses. High GPA at a strong school should warrant consideration as an exception to low GRE.

d. Ratio of national to international students. This is a reasonable surrogate for experience, and important in considering access to external support, which has a direct correlation with quality and ability to improve the overall program. The quality of institution and previous experience of international students are critical in determining quality of the student, as is performance on standardized tests, which should have a minimal threshold. A defined list of international schools should be identified so that any exceptions are examined carefully.

e. Ratios. These can be used as excellent relative measures of selectivity and quality, but are easily confounded by other factors, especially national vs. international students, and must be put in the context of absolute measures of quality. All ratios should be calculated separately for national and international students. Increasing the number of good applicants is desirable, whereas increasing the number of unqualified applicants is undesirable, irrespective of ratios.

i. Applicants to total number. High is better, and represents the ‘percent market share’ being seen by OSU. However, this can be confounded by ease of application process and marketing to increase or decrease number of applicants. This is especially true for international students.

ii. Admits to applicants. Low is better, but again can be confounded by ease of application.

iii. Enrolled to admits. High is better, but can be confounded by factors beyond strength of program, such as geography and available financial support.

II Time to Degree and Graduation Percentage

- a. Dependent upon program; care to compare OSU units to University Peer Aspirational Institutions
- b. Distribution (median vs. mean and higher moments) more meaningful than one number

Time-to-degree varies greatly by discipline, but there are well established national norms for the various fields. Performance of programs should be evaluated against these norms and specifically against the aspirational peer universities. In addition to being a quality issue, time-to-degree is also a cost issue, as greater institutional resources are invested in students who take longer to complete their degrees.

Successful graduate programs will have a low drop out rate, indicating that they have admitted students who have the capacity to perform well and that they have provided the time, energy, and resources

necessary for the student to succeed in the program. The numbers need to be looked at carefully since a few students who take many years to complete their degrees can skew the averages.

III Systematic Application of Standard Graduate Reports

Comparison of results across all programs to University averages on Graduate School Exams as compiled by the Graduate School.

The Graduate School requires an external member as part of the committee for the PhD candidacy exam and the Doctoral Dissertation Defense exam. The external member is selected by the Graduate School from the members of the P category graduate faculty on campus. The role of the external member is to evaluate the quality of the exam and to ensure fairness. Reports of performance of students in each program are sent quarterly to Graduate Studies Committee Chairs, Department Chairs and Deans, who also receive a report listing the members of their faculty who perform this service (and those who do not). These exams can be used to compare individual programs within colleges and across the university.

IV Percent Receiving Fellowship

- a) Only Fellowships that are competitive, non-departmental and non-College should be included. Institutional fellowships can be highly competitive within our walls (e.g., Presidential Fellowships) whereas others are competitive only within the context of program admission (e.g., training grant fellows). Students supported on, for example, start-up funds or college competitive funds are not to be considered within this metric, because such activities are captured elsewhere in our metrics.
- b) We are especially interested in students attracting nationally-competitive fellowships. Some programs exist that span all disciplines (notably the Fulbrights), but most are restricted by discipline. The availability of Fellowships roughly follows that of external research funding, because the national granting agencies tend to provide considerable graduate fellowship support: NSF, DOD, DOE, EPA, et al. There are also prestigious fellowships awarded by nonprofits (Hughes, Sloan, Woodrow Wilson) that should be included.

Additional comments: Percentages will be low for many programs, as a function of availability and student quality. Thus application of this metric must take into consideration availability of such fellowships by discipline: Comparisons across disciplines with this metric will be error-prone. Therefore, the metric should be used primarily to compare our programs with discipline-specific aspirational peers. Finally, we must always consider percentage metrics in light of total enrollments: 50% means something quite different for an n of 2 versus n of 20.

V Training Grants

Peer-reviewed training grants, supported by federal agencies such as NSF and NIH, are additional measures of the quality of the doctoral program. Training grants are often targeted for specific areas and not available for all graduate programs. Therefore, comparisons of graduate programs should be made with departments and programs at peer institutions for which training grants are available, e.g., sciences, engineering, biomedical areas.

One example is the IGERT training grant program from NSF, which focuses on educating U.S. Ph.D. scientists, engineers, and educators with the interdisciplinary backgrounds, strong disciplinary knowledge, and technical and/or professional skills. Also, the T32 Institutional Training Grants from NIH develop or enhance research training opportunities for individuals who are training for careers in specified areas of biomedical, behavioral, and clinical research. For these types of training grants, intensive peer-reviewed processes evaluate the objectives and direction of the training program, the quality of the faculty mentors, the caliber of the students and applicant pool, the quality of the institutional training environment, and the training record of both the program and the designated faculty.

VI GTA/GRA Ratio

OSU's aspirational peers have a significantly greater proportion of their funded graduate students working as research associates as opposed to teaching and administrative associates. This difference is largely a function of the volume and size of extramural grants and the existence of a culture that expects principal investigators to fund both stipends and tuition in grant proposals. It must be noted that the use of GRAs, especially those who are externally funded, varies dramatically across disciplines. Consequently, this metric is most valuable as 1) a single aggregate for the university; to be compared against aspirational peers with the caveat that university wide totals will vary based on the ratio between heavily funded fields (such as science and engineering) and largely non-funded fields (such as the humanities) at individual universities, or 2) discipline specific data that can be compared to like disciplines within aspirational peer institutions. The metric is not as useful when comparing across disciplines within OSU. Data on the source of support for both stipend and tuition authorization further refines this metric and allows units to better track performance as it relates to research support for graduate education.

VII Strength of Faculty within Program

Generally the quality of publications and number of citations is considered a valid and widely recognized metric for judging faculty quality. The concern is to be careful in applying these metrics between disciplines that have very different cultures. For example, while science, engineering and medical research all share a culture of publishing in journals, humanities has a culture of book publishing. Thus, any kind of simplistic comparison across the 100 programs would be invalid. This is another example where it is desirable to compare those disciplines for which journal publication and citations are common, and comparing those, while singling out other disciplines for comparison of book publication, performances, etc. Therefore, this process may require comparison with like departments in OSU's aspirational peers. The issue of measuring scholarly output in disciplines with no uniformly accepted standards is tricky and deserves closer attention

Even within those disciplines for which journal publications are the norm, weight should be placed upon those journals, specific to each discipline, that have the highest impact. To first order, citations remain a reasonably reliable measure of publication impact within a field.

The extramural support is again a highly discipline-oriented metric. We should, perhaps, give some thought to measuring trends of support, rather than any absolute measure.

The use of NRC Gini coefficients is justified for both the obvious reason that the NRC uses this methodology and because it gives an indication of whether excellence is distributed across the faculty.

Documenting external and internal awards to faculty, including service on National level Boards, appears to be a cross-disciplinary, valid measure of a program's influence on the national scene.

Some care should be applied in determining the "averaging" time for determining all of the faculty metrics, for excellence is often the result of many years of scholarly pursuit, and not a year/year measure.

NB: The Committee is well aware that gathering these data and comparing across institutions may well prove to be too time consuming and costly. An alternative is to assemble various teams of faculty within OSU to evaluate the faculty in a more subjective mode: visit the program, interview the faculty, and assess those parameters that can be readily obtained. This process might allow some of the intrinsically more difficult to measure features of a program—such as teaching, program management, etc. to be captured. This has proven useful in other ratings of cross-disciplinary programs on campus.

Charges given to the Committee by Provost Snyder on June 17, 2004

1. How can we ensure that doctoral education serves the goals of the Academic Plan? What continuing procedures should be implemented to monitor the role of doctoral education at OSU?
2. Recommend a process for assessing the quality of doctoral programs and appropriate metrics. These metrics should include, but are not limited to, appropriate external rankings as well as internal procedures.
3. Recommend a sustainable funding model for graduate education that will align state subsidy with quality. Priorities for investment are a) programs that are already ranked as very good or excellent; b) additional programs that are essential^{VII} for any great public research university (whether already strong or not at OSU); and c) programs that make unique contributions to or derive unique strength from the State of Ohio.
4. To generate resources for investment, propose a set of criteria by which I could consider the following options for programs deemed as too weak to be sustained at their current level: a) eliminating programs; b) strategically reducing the size of programs; c) freezing programs at their current size; or d) merging programs.
5. Should there be university-wide criteria on funding graduate research associates from grants? If so, recommend appropriate criteria.

* The Committee suggests that the word “essential” in this Charge be interpreted as “valuable”. “Essential” can be argued in many dimensions, and is open to gaming. “Valuable” connotes a measure of importance to the institution’s mission.