



Report of Committee on Graduate Education

Part II-Assessing Quality Of PhD Programs

Associate Provost & Interim Dean C. Anderson

Dean D. Andrews

Dean R. Brueggemeier

Ms. J. Carpenter-Hubin

Dean D. Evans

Ms. L. Flesch

Dean R. Freeman (Chair)

Dean J. Herbers

Ms. M. Mead

Dean F. Sanfilippo

EXECUTIVE SUMMARY

This Report is part II of the final report on graduate education as requested by the Provost July, 2004. Part I, submitted and approved by the Provost in September 2005, addressed parts 1, 3, and 5 of our committee's charge, focusing primarily on fiscal support of quality PhD education. Part II addresses parts 2 and 4 of the Provost's charge (see Appendix I for the Committee Charge). We broke our work up into three segments: (1) The generation and subsequent campus-wide discussion of metrics of quality to apply to our PhD programs; (2) an analysis of data held centrally concerning all of the chosen metrics, and discussions with either Department Chairs or graduate program Chairs; and (3) a detailed test of the application of these metrics on a group of 13 PhD programs.

The results of the three segments of our work are:

1. The quality metrics widely discussed across the University, and used in the test program are given in Appendix II. Why the metrics were chosen and how they may best be applied are noted in the extensive footnotes to each metric.
2. We encountered wide-spread skepticism within nearly all graduate programs that any comparison of programs that relied solely upon intra-university comparison of programs would be either accurate or useful:
 - while the metrics were applauded in general as having relevance to quality indicators, most all programs found fault with the application of one or more of them to themselves.
 - a. There was considerable discomfort expressed by many non-science programs that the metrics were not sensitive enough to disciplines that weren't easily subjected to numerical metrics;
 - b. Considerable controversy surrounded the committee's work on metrics because of the lack of faculty on the committee, and because no Dean from Arts or Humanities served on the committee.
 - much of the data necessary to field the metrics were either
 - a. not kept in an addressable format by the university and/or the programs studied;
 - b. or, when data were available from both the university and the program, they did not agree, which confounded the process;
 - virtually all of the data held by the University (either in OAA or the Graduate School) were not cleanly delineated as to the target degree: that is, there is an inherent confounding of terminal master degree students and PhD students so that the central data available were of no use
3. The application of the quality metrics to the 13 programs yielded suggestions of correlations between metrics, yet were judged to have enough program-specific differences to rule out a simple application of the metrics to all PhD programs for intramural rankings. However, our test demonstrated that these same metrics could find immediate application in alerting the Graduate School to apparently weaker programs in need of closer examination, and could be useful to Colleges for analyses and preparation for intermural comparisons in the future.

The Committee agrees with the stated intention of the Provost: the Graduate School must hold the approximately 100 PhD programs accountable for their educational quality and reputation, with rewards for excellence and substantial disincentives for continuing poor performance.

While we conclude that a university-wide application of our committee's metrics of program quality are not by themselves sufficient to determine a reliable rating of quality of programs in general, we did find that those metrics relating to "outcomes" were strongly suggestive of an enormous disparity in the quality of PhD programs at OSU. The Committee recommends aggressive measures to rank PhD programs through inter-university comparisons, and for the Graduate School to participate in and oversee external reviews of all PhD programs

We recommend three immediate actions for the Graduate School:

- *Implement immediate action to separate PhD students from Tagged or terminal Masters Programs in all data systems;*
- *Require all PhD programs to submit documentation of placement of their doctoral graduates, both for the first position of the graduate and for subsequent positions as well;*
- *Require all PhD programs to complete surveys similar to the 2006 National Research Council's program questionnaire*

Part II-Assessing Quality Of PhD Programs

A Time to Pay Attention to Graduate Programs

There are several facts concerning graduate education at OSU that make it imperative that, as an institution, OSU must pay the same level of attention to its graduate programs as it has to its undergraduate experience and quality of student. The foremost reason is that the “reputational” portion of any overall ranking statistic (e.g., US News and World Report) comes from the perception of the quality of our research programs by our peers. While our institutional efforts to increase the quality of undergraduates admitted to OSU have been widely successful, to the extent that the average admitted student has an ACT score at nearly the 90th percentile, our graduate programs have admitted students with an average GRE score which places our programs no better than in the 55th percentile.

There is a remarkable irony in such a low average GRE percentile of all OSU graduate programs when one considers that within the Arts and Sciences alone, four programs are rated in the top 25. The low average arises from the combined effect of a series of decisions and actions (or non-actions) taken by the Graduate School and/or the University as a whole over the last 15 years. Arguably the most serious is the continuing adherence to a financial model for graduate programs that rewards sheer numbers of graduate students over quality (of either the programs or students). When combined with a failure to systematically review and oversee graduate programs from a central authority, the result is a set of graduate programs that dramatically reduce the resources needed to promote OSU’s best programs to the next plateau of excellence. These financial issues are addressed in detail in Part I of this report.

Unfortunately, behavior patterns within the Colleges of the University help perpetuate our problems. Figure I documents the recent complaint¹ of the current interim Dean of the Graduate School that trying to enforce even the minimal standards of a 3.0

¹ Interim Dean Carole Anderson; Council of Deans, March 2006.

undergraduate grade point for admission to an OSU graduate program is often defeated by the Deans and Chairs of programs themselves. Without doubt this behavior is driven largely by the financial incentives which reward numbers of students rather than quality, but the words of Cassius come to mind: "The fault, dear Brutus, is not in our stars, but in ourselves...."

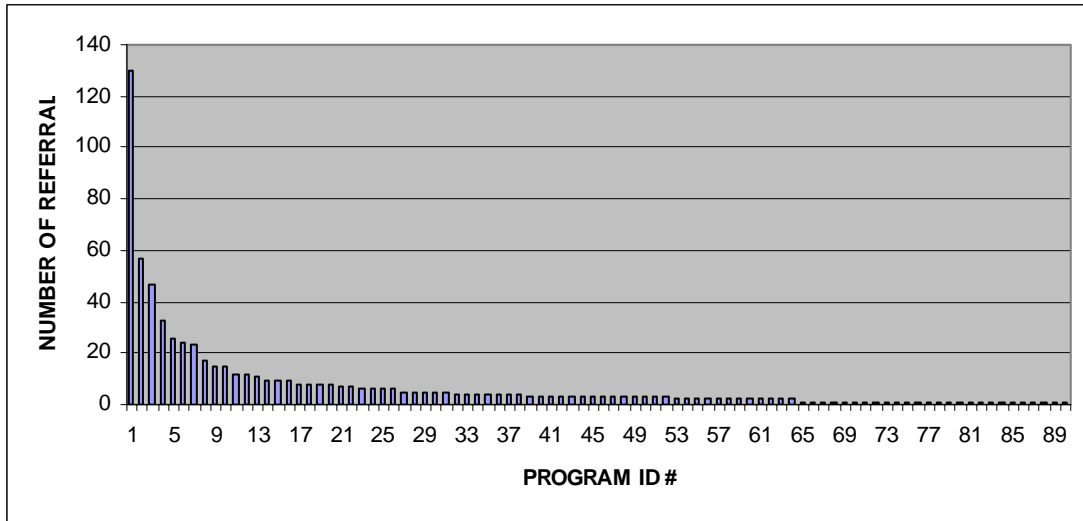


Figure 1 The number of requests granted from AU03 through SP05 for an exemption from the minimum undergraduate grade point requirement of 3.0 by program (names of programs are suppressed here).

As discussed at length in Part I of this report, there is an apparent correlation between programs in the bottom quartile of figure 1 and a figure in Part I that plots the ratio of admitted students to PhDs granted over an extended period. Although the data in that figure suffer from the admixture of Terminal Masters with PhD students -despite attempts to disaggregate the two sets- the suggestion remains that a non-negligible number of programs at OSU are seriously underperforming and systematically robbing resources from the upper quartile of graduate programs.

Design and Application of Internal Metrics of Program Quality:

For Part II, the Committee spent the majority of its time identifying the largest set of metrics that the Graduate School could reasonably apply 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 Committee noted that given the time and monetary restraints imposed upon this proposed analysis, the best result that could be expected from an application of the metrics across the university would be a broad grouping of programs into bands. The top band would include programs that should be encouraged to continue in their drive for national and international recognition. The middle band would be those programs that are either too new to OSU to be rated, are of value to OSU's Land Grant mission, 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 either (a) are not of special value to OSU's Land Grant educational mission, (b) have been historically marginal with little or no improvement, or (c) are not essential to OSU's future in terms of the Academic Plan.

The committee adopted the position that a satisfactory response to the 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 relatively short period of time. The committee 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.

1. Within the past year, Vice Provost Randy Smith announced an external review plan for systematic reviews of all academic programs at a rate of 5-10/year. The Committee is explicitly supportive of this initiative (and so calls out in Part I of this report); the goals of this work on metrics was to determine if there were internal processes that could yield the broad-banding of programs discussed above on a faster time line.

2. The generally accepted method of graduate program review is to ask 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 Northwestern University. Ohio State's program review process is modeled after Northwestern's, and while capable of fully responding to the Committee's Charges if implemented across the University as a whole will take considerable time and effort to implement, as well as substantial funding.)

3. The Committee discussed at length seven categories of data to be gathered either from university sources or the programs themselves to construct a reasonable model to analyze the PhD programs over 3-4 months. We added three supplemental metrics which, while clearly not applicable widely across all programs, could be valuable metrics when compared to the University's aspirational peers. Even within the "core" seven metrics, allowances for differences in the relative importance of the metrics by discipline were expected to be substantial. The proposed data collection for each program was designed to be consistent with the National Research Council (NRC) data collection for its upcoming study. The committee recognized immediately that the most meaningful comparisons of OSU's PhD programs would arise from inter-university, program by program analysis with the University's aspirational peers. The set of seven core and three supplemental metrics are²
 - a. Core
 - i. Entering student quality: GRE scores, quality of UG institution, undergrad GPA, ratio of national/international students admitted/enrolled, ratios of applicants to total grad student number, admits to applicants, enrolled to admits
 - ii. Time to degree and graduation rate

² Detailed analysis of the metrics are in the Appendix

- iii. Reports of the external committee members for Graduate School exams
 - iv. Percent of students receiving University and national fellowships
 - v. Success in obtaining training grants
 - vi. Ratio of GTAs to GRAs within a program
 - vii. Faculty quality: publications and journal quality, citations, extramural support, grad student/faculty ratio, awards and honors, number of associate professors and years in rank
 - b. Supplemental
 - i. Student professional activity: presentations, performances, papers, grants
 - ii. Placement
 - iii. Uniqueness of program: number of similar programs in peers, world
4. The committee acknowledged that while compiling a list of metrics to use in measuring the strength of a program is relatively straightforward, actually implementing the data-gathering process in a reasonable time and with affordable effort may be a more challenging task. The committee decided to gather data on a small subset of programs to test the feasibility of the process. The plan was to construct a model that assigned programs to the three bands and assess the model's success according to our a priori assessments of program quality. This test was designed to identify which metrics are either redundant or of little actual use in the determination of program quality.

Application of Metrics to 13 Programs:

The committee chose a group of pilot doctoral programs for which we would collect metrics data: Physics, Geological Sciences, Linguistics, Greek and Latin, Psychology, Anthropology, OSU Nutrition, Art Education, Theatre, Entomology, Microbiology, Pharmacy, and the Integrated Biomedical Graduate Program. These programs were volunteered by the Deans on the committee and augmented by programs from Arts and Sciences that were specifically chosen to have rather different objectives in PhD education than the sciences. In our initial phase, we explored in a general way the applicability of the metrics to these programs by interviewing most of the Chairs of the programs, and asking for advice. We quickly discovered that much of the data required

to construct comparisons of the metrics did not exist in the University data sources, Graduate School data sources, or within the programs. Further, any attempt to assemble these data would have to rely very heavily on the subjective judgment of the individual programs, in fact to such an extent that the Chairs warned us that whatever conclusions we would draw from the data would lack credibility with the faculty. (Perhaps not surprisingly, faculty publication data was cited most often by chairs in this regard). In addition, it was clear that many of the data would be meaningful only when compared across institutions for one discipline, not when compared one discipline to another at OSU.³

The pilot study commenced in late winter quarter, 2005 and extended through early fall quarter 2005. Institutional Research and Planning (IRP) collected data from central sources on doctoral student quality (GREs, admit and yield rates), time to degree and completion rates, and ratios of doctoral students to faculty and to graduate faculty. The Graduate School provided reports of the external committee members for Graduate School exams. A survey was conducted of pilot programs to collect additional data on graduate student support, professional activity, and placement.

Perhaps the most disappointing portion of the study was the inability to accumulate information concerning faculty productivity, either in publications, citations, or other scholarly work. We found that this material, while presumably kept in departments in some manner for P&T considerations, as well as yearly evaluations, was not in a retrievable nor common format. When we investigated computer programs that are being developed to retrieve this data from the national databases, we found them to be impractical to use. The committee found the systematic lack of basic scholarly records of department's research faculty to be disturbing, and yet another indication of the lack of attention on graduate education as related to research. The committee recommends that the Graduate School and OAA require such records to be constructed and updated in a format that can be used for accreditation, data exchange, and internal rankings.

³ The committee concluded that, with few exceptions, the only data that would be broadly accepted across the University as meaningful comparisons would have to be inter-university, program to program. This is discussed in more detail below.

In Figure 2 is a summary of the IRP available data that were shared with each program and compared to the program data for the 13 pilot programs. In each case, considerable time and effort was expended in comparing and qualifying data with the programs. The committee was not surprised to find that centrally held data were, in many cases, not consistent with program held data. In addition, we accumulated data on placement of doctoral graduates (to the extent the programs had the data). We met with each of the Chairs of the programs and discussed the data in Figure 2 and the placement data. The data in Figure 2 have the names of the programs suppressed, and the columns listed in arbitrary order. Its purpose for this report is not to (perhaps inappropriately) label any program as underperforming, but to demonstrate the large dispersion in program numbers. In this figure, each column is one of the pilot programs; dark shaded cells are nominally at the higher end of the test programs' distributions, while light shaded cells are nominally at the lower end. What is evident is that there are programs (columns) that have a large proportion of heavy shaded regions, while there are others with a large proportion of lightly shaded regions.

GRE SCORES - 5-yr averages (2000-2004)													
Applicants Average Verbal GRE	565.2	524.3	600.6	523.5	569.7	555.3	534.8	452.9	511.1	548.4	561.3	560.0	528.6
Applicants Average Quantitative GRE	728.9	688.8	600.9	585.7	694.0	774.2	625.9	550.9	653.6	569.4	708.9	739.6	671.9
Applicants Average Analytic GRE	674.8	652.2	651.2	619.0	629.7	701.5	641.7	504.3	617.8	599.4	676.0	695.0	660.6
Applicants Average Writing GRE	4.1	4.3	5.2	4.7	4.6	4.1	4.8	3.9	4.1	4.7	5.0	4.3	4.5
Admitted Students Average Verbal GRE	552.8	517.5	609.4	544.1	657.1	541.1	611.8	477.8	517.4	588.4	530.3	535.7	545.4
Admitted Students Average Quantitative GRE	681.7	665.4	606.4	598.2	724.2	748.3	691.6	552.2	659.8	572.6	652.4	707.7	676.0
Admitted Students Average Analytic GRE	660.9	644.3	664.5	656.7	730.0	698.8	700.3	512.2	595.0	621.9	672.2	680.1	688.3
Admitted Students Average Writing GRE	4.5	4.6	5.3	4.9	5.7	4.8	5.0	4.1	4.4	4.8	5.0	4.4	4.6
Enrolled Students Average Verbal GRE	573.3	504.8	586.3	539.0	640.2	518.0	597.7	470.0	492.6	568.3	551.3	528.9	534.7
Enrolled Students Average Quantitative GRE	677.3	659.6	595.4	582.4	720.6	736.2	680.9	542.5	643.3	566.7	635.4	697.9	660.6
Enrolled Students Average Analytic GRE	644.6	618.5	667.0	655.9	718.5	686.4	688.7	501.8	595.8	633.3	657.0	669.8	674.3
Enrolled Students Average Writing GRE	4.1	4.5	5.2	5.0	5.6	4.8	5.0	4.0	3.6	4.9	5.0	4.3	4.6
GRE SCORES - change in 3-yr rolling averages (2000-2002 vs 2002-2004)													
Applicants Average Verbal GRE	14.8	15.8	2.6	-4.6	29.0	36.9	0.4	7.1	19.5	-2.7	-69.1	9.1	2.9
Applicants Average Quantitative GRE	-4.4	21.5	13.1	-2.8	18.6	6.5	11.2	18.3	27.4	-5.2	-45.1	3.0	-25.2
Applicants Average Analytic GRE	33.1	32.9	15.4	8.4	29.5	35.0	16.3	18.5	39.8	5.8	-33.8	24.9	7.6
Applicants Average Writing GRE													
Admitted Students Average Verbal GRE	7.2	-16.5	30.4	-16.5	6.5	85.8	9.7	18.0	6.7	-11.7	-13.2	-15.1	-4.2
Admitted Students Average Quantitative GRE	-25.3	-5.0	1.4	8.4	0.0	17.9	16.5	26.8	19.7	-5.1	31.3	-12.6	0.7
Admitted Students Average Analytic GRE	-0.7	2.9	46.4	-0.4	-16.3	24.4	5.5	42.4	43.4	-3.8	-11.1	15.3	3.5
Admitted Students Average Writing GRE													
Enrolled Students Average Verbal GRE	-21.4	-28.8	62.2	-18.6	10.5	79.8	11.9	17.4	-55.4	-26.0	10.0	-9.2	0.1
Enrolled Students Average Quantitative GRE	-16.9	-8.9	-10.6	16.4	-2.9	8.7	4.3	31.9	26.2	-26.2	-29.0	-19.6	-0.9
Enrolled Students Average Analytic GRE	-32.6	-15.0	39.5	-3.8	-0.8	2.5	-7.7	45.8	37.9	-8.9	NA	13.3	-16.8
Enrolled Students Average Writing GRE													
Diff OSU vs National Avg for all Test-takers by Discipline 2003-04													
Verbal	-19	-15	NA	-33	11	-14	21	NA	5	-10	TBD	13	-22
Quantitative	13	9	NA	-13	10	-2	46	NA	1	7	TBD	6	8
Analytical Writing	1	0	NA	0	0	0	0	NA	1	0	TBD	0	1
Admission/Enrollment - 5-yr averages (2000-2004)													
Number of International Students Admitted	4	2	1	2	5	7	10	5	3	1	0.4	11	5
Number of Domestic Students Admitted	2	17	8	24	8	12	61	31	8	19	4	16	20
Number of International Students Enrolled	2	1	1	0.4	3	4	5	3	2	0.2	0.0	7	3
Number of Domestic Students Enrolled	1	10	5	9	3	8	27	26	3	10	2	11	7
Ratio of Domestic to International Students Admitted	0.7	8.6	10.5	15.0	1.7	1.8	6.0	6.5	3.0	15.8	11.0	1.4	4.0
Ratio of Domestic to International Students Enrolled	0.8	9.8	6.3	22.0	1.3	2.1	5.0	7.8	1.8	52.0	NA	1.5	2.5
Percent of applicants admitted	23.8%	37.6%	40.0%	38.3%	14.4%	6.6%	15.0%	57.1%	43.0%	27.4%	14.0%	13.3%	27.5%
Percent of admits who enroll	51.6%	56.3%	69.6%	37.5%	46.3%	65.2%	45.3%	82.8%	57.7%	59.4%	41.7%	64.7%	41.6%

Figure 2: Programs are in columns (names suppressed). Heavy shaded regions are nominally at the higher end of the test programs' distributions, while light shaded cells are nominally at the lower end

To many members of the committee, Figure 2 seemed *prima facie* evidence of program quality. Upon deliberation, however, the committee decided that there were program specific qualifications that could account for much of the dispersion. Still, the wide range of fundamental measures of student preparedness and the wide disparity in the ratio of admissions to applications certainly raise questions of whether enough program oversight or control has been exercised at OSU.

We found some of our proposed metrics to be essentially useless:

Graduate School External Reviewer reports (almost no variation)

Training Grants (available to a limited number of programs)

Time to Degree (data for programs extremely chaotic from year to year)

Student Professional Activities (largely not compiled by programs)

In our discussions with the pilot program chairs we found wide-ranging agreement on one metric that appears to have great usefulness for judging program quality independent of intermural program-to-program comparison: **Placement of Graduates**. The committee examined placement data, and with some caveats, concluded that this metric was largely independent of program-specifics, with the added merit of being the ultimate “outcome” measure of a program. This recommendation was made to the Provost and Graduate School in November of 2005, and data collection for all of OSU’s PhD programs is now being collected.

In order to make the placement data relatively immune to program manipulation, several issues needed to be addresses. The first issue is how to rate the quality of placement; for example:

- is a given industrial position equivalent to a professorship at a given level of university;
- is a postdoctoral position at a national laboratory, or at a prestigious institute equivalent to a full time position at a less well known institution

In other words, since placement quality is by its very nature program-specific, how does the University determine what constitutes high quality placement for a program?

The committee has agreed upon the following process:

- Require each program to respond to the request for graduate placement, both initial and subsequent employment.
- Require each program to rate all placements into at least three categories: high quality, average, and low prestige, based upon general guidelines promulgated by the Graduate School.
- Require each program to send its rankings of placement positions to at least 5 similar departments within the University's aspirational peer list, as well as the CIC, asking for comment on rankings of placement positions.
- Require each program to reconcile its rankings with comments and counter rankings of the programs contacted.
- Have the Graduate School assign all placements by all programs into the three levels above.

With this data for each program, the University can begin sorting its PhD programs into the three broad categories outlined in the committee's original agenda. This conclusion is not to reduce the other strong recommendations of the committee, specifically that:

- external program reviews, now beginning, be accelerated and put on a continuing basis;
- that all programs be required to complete a survey similar to the NRC program questionnaire in preparation for data exchange and for institutional accreditation;
- that as soon as possible, IRP institute data exchange, program by program with the AAU universities.

Finally, the Committee is unanimous in encouraging the Graduate School, under the leadership of Interim Dean Anderson, to consider possible funding models that address the underlying reason for the misalignment of graduate education resources with excellence. The importance of attracting a highly qualified scholar and manager to be the next Dean of the Graduate School cannot be over emphasized.

APPENDICES:

Charge to the Committee

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 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.

Metrics:

- I. Judging the quality of entering graduate student within programⁱ:
 - 1) Primary indicator is GRE scores (both general and subject specific, if offered)
 - 2) Quality of UG institution (as roughly determined by USNEWS)
 - 3) Undergraduate GPA (possibly normalized by approximate- within 25%-USNEWS ranking of undergraduate school)
 - 4) Ratio of national to international students admitted and/or enrolled compared with similar programs at our aspirational peers
 - 5) 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)
 - iii. What is the ratio of enrolled/admits for the program (high is better)
- II. Time to Degree and Graduation %ⁱⁱ
 - 1) Vary, dependent upon program; care to compare OSU units to University Peer Aspirational Institutions
 - 2) Distribution (median vs. mean and higher moments) more meaningful than one number
 - 3) Master required/yes/no (separate programs in the analysis?)
- III. Systematic Application of Standard Graduate Reportsⁱⁱⁱ
 - 1) 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}
 - 1) Only Fellowships to count are competitive, non-departmental, non-College.
 - 2) 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
 - 1) Applicable only to programs eligible (compare University Peer Aspirational Institutions)
 - 2) Historical as well as current success in obtaining Training Grants

- VI. Ratio of GTA/GRA within program^{vi}
 - 1) Highly dependent upon program; meaningful comparison only by using University Peer Aspirational Institutions Data
 - 2) % 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 waiver

- VII. Faculty Quality Indicators^{vii}
 - 1) Use of NRC Gini Coefficients to measure:
 - i. Publications per graduate faculty
 - 1) Quality of Journals
 - ii. Citations per graduate faculty
 - iii. Extramural support per graduate faculty
 - iv. Graduate Student/faculty ratio
 - Distribution of faculty who actively supervise graduate students
 - 2) % 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
 - 4) University wide honorifics:
 - a) Distinguished Scholar
 - b) Distinguished Teacher
 - c) Distinguished University Professor
 - 5) College-wide honorifics
 - a) College Distinguished Professor
 - 3) Number of Associate Professors and years in rank

SUPPLEMENTAL METRICS

- 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

1 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 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 correlate 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.

1 Time to Degree and Graduation Percentage

- a. Very 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.

1 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.

1 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) 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.

1 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.

1 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 waiver further refines this metric and allows units to better track performance as it relates to research support for graduate education.

1 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 of 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. This process may, therefore, 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.

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 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 correlate 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. Very 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) 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. PhD 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 waiver 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 of 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. This process may, therefore, 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.