2009 Graduate Student Support Review Panel Final Report

December 18, 2009

The Graduate Student Support (GSS) Review Panel: Origin and Charge

In spring, 2009, occasioned by unanticipated budget challenges to the 2009-10 budget that were projected to carry over into 2010-11, the Provost called for a graduate student support (GSS) review of doctoral programs that would provide the opportunity for the campus to undertake a deliberative, multi-perspective evaluation of our graduate "portfolio." She explained that this review process was intended to be different from the individual program reviews undertaken by the departments in recent years (although it might well have drawn from those); rather, it was intended to look across the campus at the quality of our doctoral programs, to update our knowledge about our areas of strength, and to consider from there how we might best invest in our graduate programs going forward, particularly in the context of difficult budget decisions ahead. The initial charge to the GSS Review Panel was to focus on the doctoral level, although she did not rule out subsequently expanding the review as needed to evaluate the support portfolio at the master's level.

During the spring of 2009, the Provost:

(a) solicited names of potential members for the panel from the Governance Council of the University Senate (GOV), specifying distinguished faculty with extensive experience in doctoral education;

(b) asked the Dean of Graduate Studies and the Director of Institutional Research, Planning, and Effectiveness (IRPE) to meet with several Councils of the Senate and other stakeholders to gather input on data variables and design of the overall process; and

(c) laid out an ambitious time-line for the work of the panel, asking it to complete its deliberations and share its results with her by Thanksgiving, 2009.

Thus, in April, the Dean of Graduate Studies and the Director of IRPE met with members who had been involved in the previous doctoral review process in 1997/98 and included in that meeting the chair of the Graduate Academic Council (GAC) and the Chair of the University Senate. Following that meeting, the Dean and Director met with the Council on Research (COR), with the Deans of the Schools and Colleges, with the University Planning and Policy Council (UPPC), with the Council on Academic Assessment (CAA), with the Graduate Academic Council (GAC), and with graduate program directors and chairs of departments with graduate programs in order to solicit

input on the review design, process, and data metrics to be gathered and considered.

In late April, in consultation with the GOV chair, the Provost identified members of the GSS Review Panel, which ultimately included 16 persons whose appointments were specifically approved by GOV and another 11 selected by the Provost in consultation with the Dean of Graduate Studies so that the panel would reflect diverse disciplinary perspectives. The final panel included members from numerous departments that offer doctoral programs across the various Schools and Colleges throughout the University.

In early May, the Provost invited members of the GSS Review Panel and convened the panel for an initial meeting to learn about the origin of the review; to consider, review, and finalize process issues; and to establish a plan of work and a time-line. The timeline included IRPE data collection and summary throughout the summer months to be shared with department chairs in late August. Also included was a plan for the Dean of Graduate Studies to circulate to department chairs and doctoral program directors a description of the Program Document in July explaining the departmental data that would need to be collected and submitted in September to the GSS Review panel.

By the conclusion of the process, 35 doctoral programs were included in the GSS Review.

GSS Review Panel members:

James Acker, Distinguished Teaching Professor, School of Criminal Justice (Review Panel co-chair)
David Andersen, Distinguished Service Professor, Public Administration
Arthur Applebee, Distinguished Professor, Educational Theory and Practice
George Berg, Associate Professor, Computer Science
Eric Block, Distinguished Professor, Chemistry
Ronald Bosco, Distinguished Professor, English
James Dias, Associate Professor, Biomedical Sciences
Ricky Fortune, Professor, School of Social Welfare
Myrna Friedlander, Professor, Educational and Counseling Psychology
Jagdish Gangolly, Associate Professor, Informatics
Virginia Goatley, Associate Professor, Reading
Glyne Griffith, Associate Professor, English and Latin American, Caribbean, and U.S. Latino Studies
Richard Hamm, Professor, History
Edward Hannan, Distinguished Professor, Health Policy, Management, and Behavior, and Associate Dean, School of Public Health
Teresa Harrison, Professor, Communication
Kajal Lahiri, Distinguished Professor, Economics
Daniel Levy, Distinguished Professor, Educational Administration and Policy Studies
Carolyn MacDonald, Professor, Physics

Jonathan Mandle, Associate Professor, Philosophy
Albert Millis, Professor, Biological Sciences
Bruce Miroff, Professor, Political Science
Karin Reinhold, Associate Professor, Mathematics and Statistics
Lawrence Schell, Professor, Anthropology and
Epidemiology and Biostatistics
Christopher Thorncroft, Professor, Atmospheric and
Environmental Science
David Wagner, Professor, Sociology
Kevin Williams, Professor, Psychology (Review Panel co-chair)
David Wills, Professor, Languages, Literatures, and Cultures

Programmatic Dimensions Assessed and Rating Scales

Following consultation with members of the University community described above, including deans, department chairs, and representatives of the University Senate, the dimensions of doctoral programs to be evaluated during the GSS review process, with accompanying indicators, were identified. A six-point rating scale was adopted to assist with the assessment of each dimension. The programmatic dimensions selected for evaluation were:

1. *Faculty Quality* (How prominent have the faculty become in their research and/or scholarly contributions? How does the constituency of faculty in the department suggest promise for the future?)
2. *Student Quality* (What does the evidence demonstrate concerning the quality of doctoral students in the program? What evidence suggests involvement of doctoral students with undergraduate teaching, research contribution, or other workload, as appropriate to the discipline?)
3. *Program Efficiency/Effectiveness* (How well does the program use its faculty and student resources to insure that students successfully complete their programs in a timely manner?)
4. *Program Outcomes* (What does the evidence demonstrate concerning students graduated, opportunities available to graduates, and graduate students' achievements while in the program?)
5. *Overall Rating*

The six-point rating scale used to evaluate each dimension was as follows:

1. Distinguished
2. Excellent
3. Strong
4. Adequate
5. Marginal
6. Weak

A copy of the University at Albany GSS Doctoral Panel Review Graphic Rating Scale, including the program dimensions evaluated, relevant indicators, and the rating scale employed, is attached as follows:

University at Albany GSS Doctoral Panel Review Graphic Rating Scale

5. OVERALL RATING **Derived holistically from the four dimensions, with greater weight given Faculty Quality, but some weight given to all four dimensions.**	**Distinguished** **1** ***Best Practice – Program with national/ international standing and makes significant contributions to achieve University goals and campus productivity***	**Excellent** **2** ***Meritorious Performance – Program approaching national standing and makes significant contributions to achieve University goals and campus productivity***	**Strong** **3** ***Solid Performance – Program with regional or niche reputation and makes significant contributions to achieve University goals and campus productivity***
	Adequate **4** ***Minor problems – Program likely to see emerging reputation and increasing contributions to University goals and campus productivity***	**Marginal** **5** ***Serious problems – Requires significant changes in both faculty and/or student quality to get to Adequate and program makes relatively minor contributions to University goals and campus productivity***	**Weak** **6** ***Important areas of weakness – Any emerging reputation for program is based on a minority of faculty members; program needs to increase contributions to University goals and campus productivity in order to match Adequate programs***

THE FOUR DIMENSIONS	INDICATORS FOR EACH DIMENSION		
1. FACULTY QUALITY ***How prominent have the faculty become in their research and/or scholarly contributions? How does the constituency of faculty in the department suggest promise for the future?***	*Indicators:* *• NRC ranking if available* *• AA productivity (books, journals, federal funding, citations, honors/awards)* *• UA Research Foundation funding (where relevant)* *• Add'l info concerning faculty grants (from department)* *• Program-designed measures (from program assessment, if provided by department)* *• Domestic and international diversity* *• Representation at faculty ranks* *• Dept. responses to open- ended questions*		
2. STUDENT QUALITY ***What does the evidence demonstrate concerning the quality of doctoral students in the program? What evidence suggests involvement of doctoral students with undergraduate teaching, research contribution, or other workload, as appropriate to the discipline?***	*Indicators:* *• application, acceptance, and new student enrollment* *• UG GPA from sending institution* *• Grad GPA from sending institution (if relevant)* *• GRE Scores (consider full range of GRE data provided by IRPE)* *• TOEFL Scores (for international students)* *• Program-designed or collected measure(s)* *• Domestic and international diversity of students* *• Dept. responses to open-ended questions* *• Other information provided by department (such as typical goal for size of entering doctoral class, including goals for diverse recruitment)*		
3. PROGRAM EFFICIENCY / EFFECTIVENESS ***How well does the program use its faculty and student resources to insure that students successfully complete their programs in a timely manner?***	*Indicators:* *• Time-to-degree* *• 3rd year/5th semester retention rate (viewed in the context of available University resources allocated to program)* *• Disciplinary "norms" or benchmarks that provide context for TTD and 3rd year retention (where available and as supplied by departments with source cited)* *• Dept. responses to open-ended questions* *• Other information provided by department (such as number of doctoral students regularly supported on external grants or contracts, with sources)*		
4. PROGRAM OUTCOMES ***What does the evidence demonstrate concerning students graduated, opportunities available to graduates, and graduate students' achievements while in the program?***	*Indicators:* *• # of doctoral degree recipients* *• Student placements (and consistency with program goals as provided by department—e.g., # of tenure-track positions, # of post-docs, # of placements in private foundations or other organizations or other relevant measure))* *• Papers published and/or presented at conferences and awards received by students while still working toward degree (department provides information)* *• Reputational standings (US News, other published 3rd party national standings based on reputation; dept. provided, with sources)* *• Dept. responses to open-ended questions* *• Other information provided by department*		

Preparation and Submission of Program Documents

Over the course of the summer, directors of each of the doctoral programs evaluated were given detailed instructions for preparing a Program Document to be used in the GSS review process. They also were provided with relevant data assembled by the UAlbany Office of Institutional Research, Planning, and Effectiveness (IRPE) and by Academic Analytics (AA), a private company with the "goal . . . to provide universities with an annual release of accurate data on faculty performance in a comparative, disciplinary context." The Program Documents, comprising narratives addressing these data and describing relevant aspects of the individual doctoral programs, were submitted with the accompanying IRPE and AA data to Dean of Graduate Studies Marjorie Pryse in mid-September for circulation to the sub-panelists assigned to review each program.

The Formation of Three-Member Sub-panels and Their Program Reviews

In late August, prior to the submission of departmental Program Documents, the GSS Review Panel received its charge from Provost Phillips. Each panel member thereafter was assigned to one of nine three-person "sub-panels" and one of two "sub-groups." The sub-panels were created and programs assigned to the sub-panels based on the following criterion: no sub-panel could review a doctoral program that included as one of its members a faculty member in that program or one who enjoyed a joint or affiliate appointment to that program. In addition, the co-chairs were careful to exclude any program from consideration by a specific sub-panel where one or more members of the sub-panel might directly or indirectly benefit from a higher rating assigned to that program. The sub-groups emerged because it proved impossible to find a single meeting time suitable for the entire 27-member GSS Review Panel, and the co-chairs saw the need to meet with the Panel as a group on several occasions throughout the process. However, both the introductory and concluding meetings of the Panel were arranged at a time when the entire group could meet as a single body.

In preparation for reviewing the UAlbany Program Documents, each of the three-member sub-panels was provided with mock Program Documents attributed to two fictional departments (Botany and Scandinavian Studies). The sub-panels evaluated the hypothetical programs, employing the six-point rating scale on each of the dimensions assessed (Faculty Quality, Student Quality, Program Efficiency/Effectiveness, Program Outcomes, and Overall). Each sub-panel produced a Program Review in explanation of its ratings for each of the fictional programs. The Program Reviews prepared by the respective sub-panels were distributed among all GSS Review Panel members. The GSS Review Panel subgroups met and used the Program Reviews to engage in a "norming" exercise designed to help ensure consistency among the sub-panels in their use of the assessment criteria and rating scales. In addition, the Dean of Graduate Studies invited a faculty researcher from the Center for Policy Research who specializes in group decision-making to attend and observe the "norming" exercise. The Dean and Director of IRPE had informally consulted with him during the summer, so he was familiar with the GSS review process. When he attended the GSS panel "norming" exercise, he had received copies of the fictional program documents, as well as documents related to the GSS

Review design, and later shared his observations with the co-chairs. He encouraged the co-chairs to believe that the design would produce adequate inter-rater reliability and he also made suggestions for resolving significant differences in ratings, should those occur, and for sharing responsibility for leadership within the sub-panels.

Between mid-September and early November, each three-member sub-panel was asked to review and evaluate the Program Documents of (typically) seven or eight UAlbany doctoral programs. As noted above, no sub-panel member was affiliated with any of the programs evaluated by his or her sub-panel. Two sub-panels, working independently of each other, evaluated each of the 35 doctoral programs assessed. A strict confidentiality policy was maintained throughout the process, and other than the GSS Review Panel co-chairs, no sub-panel members knew which programs were evaluated by other sub-panels or even which panelists belonged to particular sub-panels (although occasionally someone would identify a co-sub-panelist in general discussion). As questions arose or as additional information was considered necessary during the review process, the sub-panels were asked to contact one of the Panel co-chairs who, in turn, contacted the department whose Program Document was at issue in order to secure the requested information.

The sub-panels' evaluations were made in the form of Program Reviews that included a numerical rating (1-6) for each dimension of the doctoral program evaluated (Faculty Quality, Student Quality, Program Efficiency/Effectiveness, Program Outcomes, and Overall) and an accompanying narrative in explanation of the ratings. Program Reviews typically were two or three pages, although some were a bit shorter and some a bit longer. The sub-panels' Program Reviews were submitted to the GSS Review Panel co-chairs on or before November 4.

Preliminary Analysis and Discussion of Program Reviews

The Panel co-chairs made a preliminary review of the sub-panels' Program Reviews. One sub-panel was asked to elaborate more fully in explanation of the ratings assigned to some of the programs it reviewed, and those fuller reports were provided. One sub-panel was asked to reconsider its assessment of one program, when the Overall numerical rating did not correlate with the numerical ratings of the underlying four programmatic dimensions. The involved sub-panel reconvened and submitted a revised Program Review. Analyses were conducted of the numerical ratings assigned by the sub-panels, including measures of inter-rater reliability for the two sub-panels that reviewed the same program. A summary of those analyses is attached. A comparison of the numerical ratings assigned to the "Overall" program assessment by the two sub-panels that independently evaluated each of the 35 doctoral programs reflected the following:

Identical Overall Rating	13 programs
One-point Differential	19 programs
Two-point Differential	2 programs
Rating Made by Only One Sub-panel	1 program

The GSS Review Panel sub-groups met following submission of the sub-panels' Program Reviews to discuss next steps and the general results of the assessments, including the analyses of the numerical ratings. Programs were not identified by name, nor was it revealed which sub-panels evaluated which programs. Following discussion, the Panel decided that in all cases in which the two sub-panels evaluating a program reported identical Overall ratings or in which a one-point differential existed, the Program Reviews prepared by each sub-panel would be transmitted to the director of the program evaluated without reconvening the sub-panels. For the two cases in which sub-panels evaluating a program differed by two points in their Overall numerical rating, the co-chairs notified the involved sub-panels about the two-point discrepancy and each was invited to reconsider its original assessment in light of knowledge of the discrepancy and to make changes or retain the original assessment, as it deemed appropriate. It further was decided that if any changes were made in Program Reviews in such cases, the director of the involved program would be provided with copies of both the original assessment and the modified assessment, with an explanation of the underlying process. Following this process of re-examination, the discrepancy in Overall ratings was reduced to one point for one of the involved program and remained at two points for the other program.

Final Analysis

The extent of agreement between sub-panels was assessed by examining absolute agreement in ratings and by calculating intra-class correlation coefficients. The following table presents the frequency with which the final sub-panel ratings agreed or varied by one point, two points, or three points. The vast majority of ratings were within one point of each other. Across all assessment dimensions, sub-panels gave identical ratings in 43% of cases (73/170 ratings) and varied by one point or less in 93% of cases (158/170 ratings). For overall ratings, sub-panel agreed on the ratings 38.2% of the time (13/34), and differed by 1 point 58.8% of the time (20/34). Sub-panels varied in their overall ratings by 2 points on just one occasion. Thus, there was considerable convergence in the overall ratings of doctoral programs, as sub-panels agreed or varied by just one point in 97.1% of case. Rating convergence was slightly lower for program efficiency and program outcome.

Frequency of Sub-Panel Agreement and Rating Differential by Assessment Dimension

Rating Dimension	**Absolute Agreement**	**1 point difference**	**2 point difference**	**3 point difference**
Faculty quality	17 (50%)	15 (44.1%)	2 (5.9%)	0 (0%)
Student quality	17 (50%)	16 (47.1%)	1 (2.9%)	0 (0%)
Program efficiency	15 (44.1%)	17 (50%)	2 (5.9%)	0 (0%)
Program outcomes	11 (32.4%)	17 (50%)	5 (14.7%)	1 (2.9%)
Overall rating	**13 (38.2%)**	**20 (58.8%)**	**1 (2.9%)**	**0 (0%)**

A more comprehensive assessment of agreement was conducted using intraclass correlation coefficients (ICC). An ICC provides a measure of the proportion of variance that is attributed to targets of measurement (doctoral programs in this case). ICC will approach 1.0 when there is no variance in program ratings provided by different sub-

panels. There are different ICCs for different measurement designs. In the current design, programs were rated by different sets of sub-panels. That is, the sub-panels who assessed one program were not necessarily the same as those who assessed another. The appropriate ICC for this design corresponds to a one-way analysis of variance (ANOVA) design in which program is a random effect, and sub-panel (rater) is treated as measurement error.[1] This type of ICC provides an estimate of the degree of absolute agreement among ratings. ICCs were computed for each of the five dimensions and are presented in Table 2. Two versions of the ICC are reported: one based on the reliability of the individual assessments provided by sub-panels ("single measure" ICC) and one based on the reliability of the average of the two sub-panel assessments ("average measures" ICC). The single measure ICC is equivalent to traditional notions of inter-rater agreement; the average measure ICC provides an estimate of agreement in the population of raters if averages of the sub-panel ratings were used.[2]

Table 2. Intraclass correlation coefficients (ICC) by Assessment Dimension

	Assessment Dimension				
ICC type	**Faculty quality**	**Student quality**	**Program Efficiency**	**Program Outcomes**	**Overall Rating**
Single measure	.64	.56	.62	.44	.61
Average measure	.78	.72	.76	.61	.76

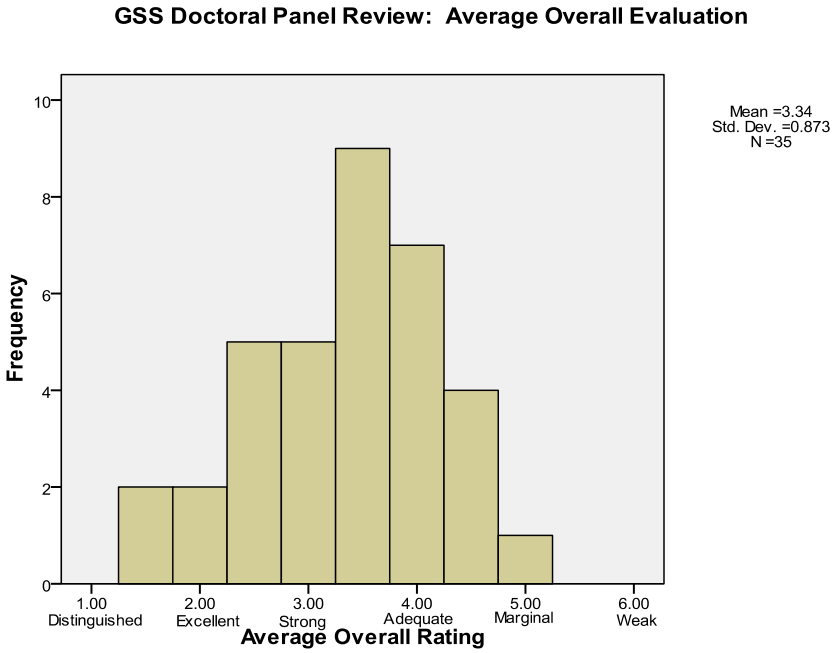
As with other measures of inter-rater reliability and agreement, there is no standard acceptable level of agreement using an ICC. Although a reliability coefficient of .70 is often recommended as a minimal standard for a measure, meta-analyses indicate that the average level of inter-rater reliability across different rating contexts is between .50 and .55. Thus, an ICC of .60 may be used as an acceptable level of agreement for the GSS design. Overall, sub-panel assessments showed moderate and acceptable levels of agreement in faculty quality, program efficiency, student quality and overall rating. Agreement was lower for program outcomes. It should be noted that program outcome ratings were based on data provided by programs. These data were not uniform across programs, which may have contributed to lower levels of agreement.

A generalizability analysis was also conducted to assess the agreement across all rating dimensions, programs, and sub-panels (i.e., the level of agreement for any single rating pulled from the program X dimension X rater matrix). The resulting ICC for a single measure was .35, which indicates that 35% of the total variance associated with a single rating from a sub-panel can be attributed to differences between doctoral programs. The ICC for ratings averaged across 2 sub-panels and 5 rating dimensions was .70, which indicates an acceptable level of reliability for the average of sub-panel ratings.

[1] This corresponds to "Case 1" or "ICC(1)" in the psychometric literature [McGraw, K. O., & Wong, S. P (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30-46; Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428]. It provides an estimate of the degree of absolute agreement among measurements of different objects.

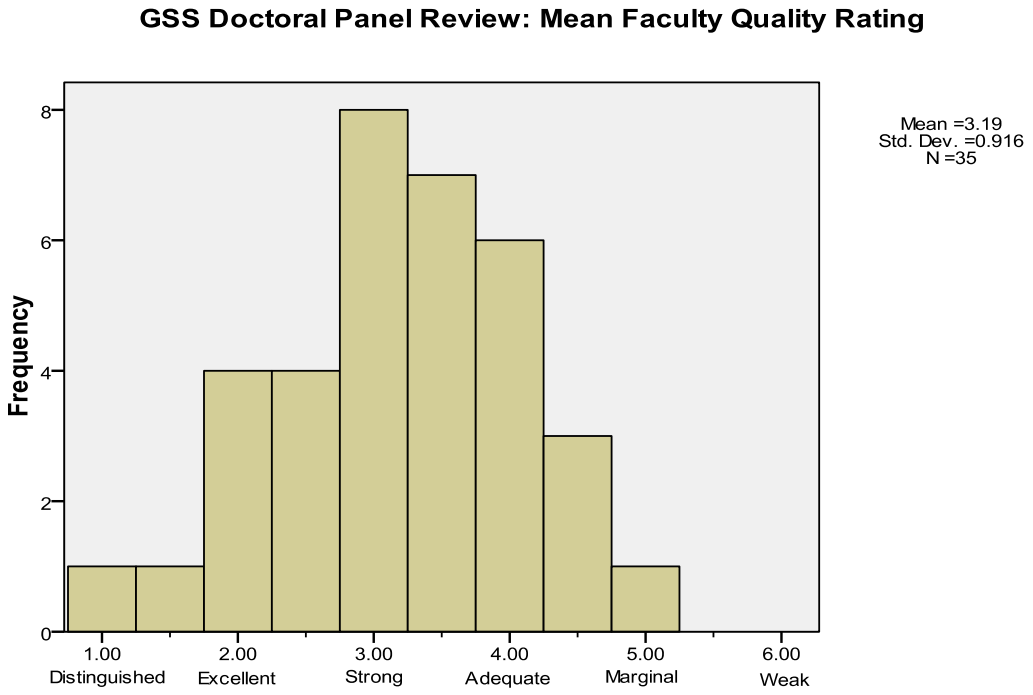
[2] From a practical viewpoint, the single measure ICC is appropriate if individual subpanel ratings are retained and treated separately; if the average of the subpanel ratings is computed, then the average measure ICC is appropriate.

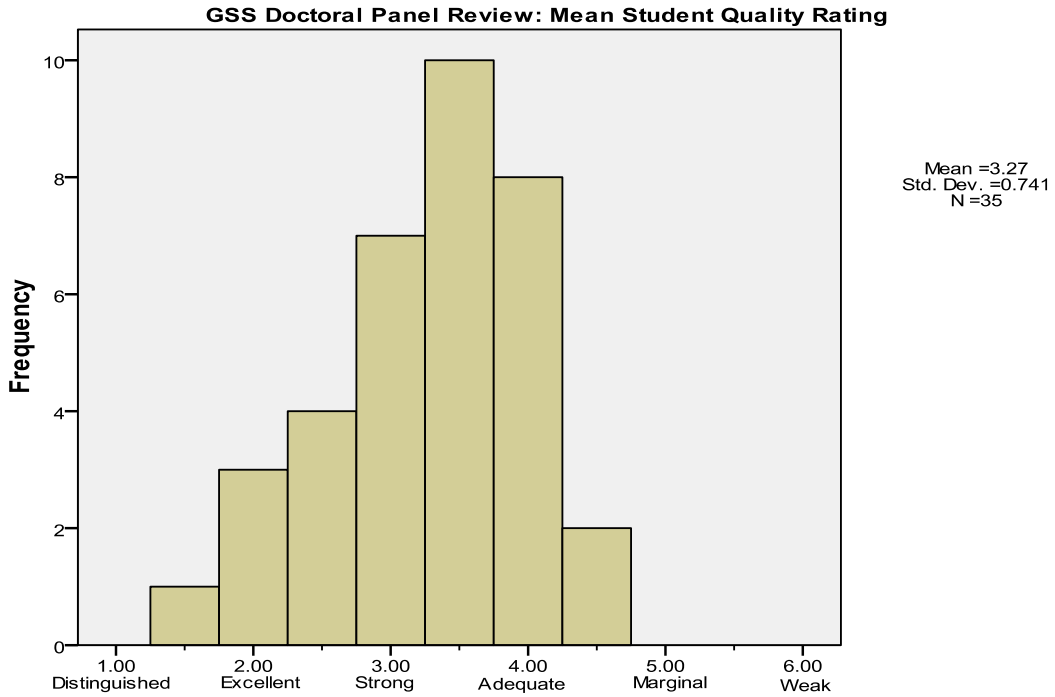
Frequency distributions and descriptive statistics were analyzed for the five rating dimensions. The frequency distribution for the mean of the sub-panels' overall ratings is displayed in the following histogram.

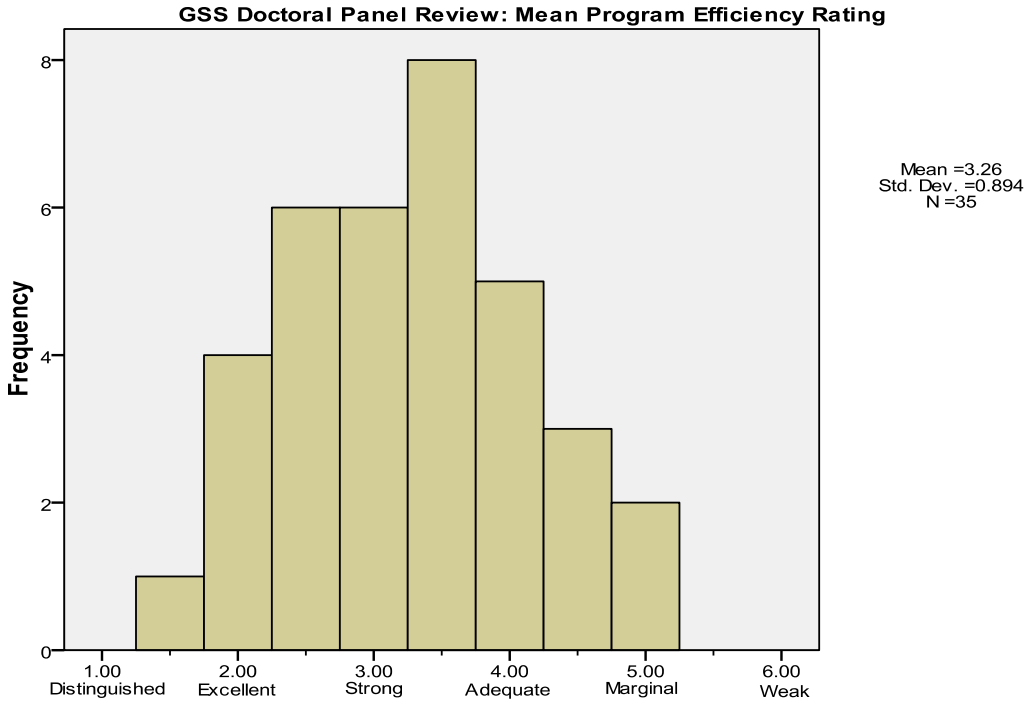


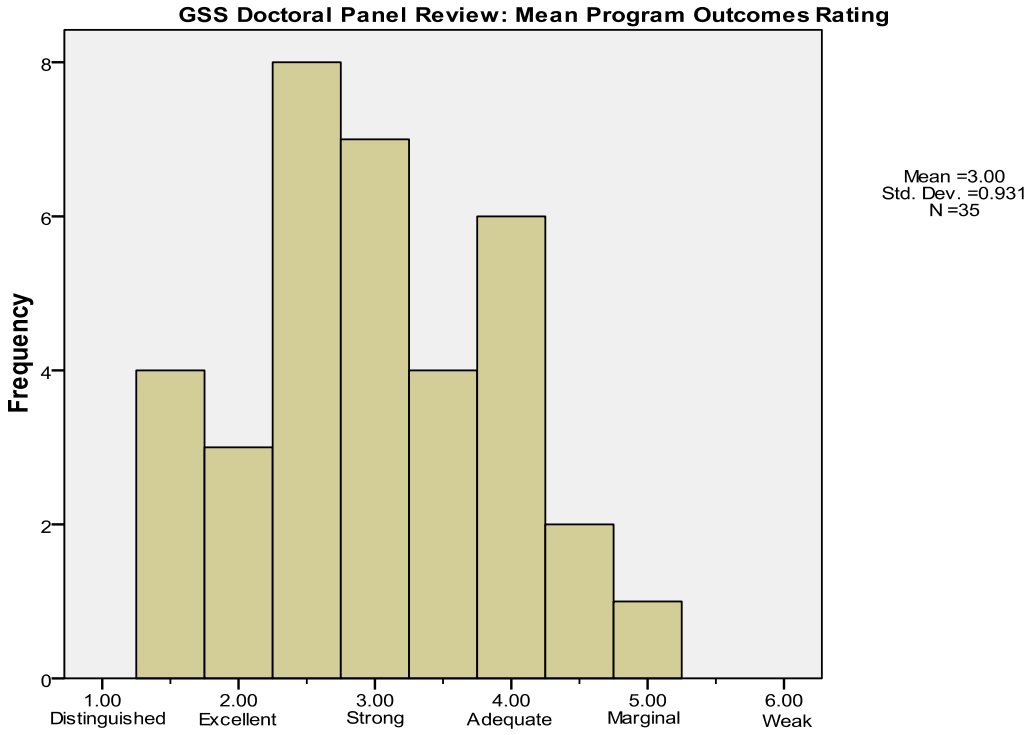
The average of the sub-panels' overall ratings ranged from 1.5 to 5 on the 6-point rating scale. The modal rating was 3.5 and the mean was 3.34. The majority of programs were clustered between "Strong" and "Adequate," yielding an approximation of a bell-shaped curve rather than a uniform or flat distribution. The distribution displays a slight negative skew (-.37), indicating that the ratings are clustering toward the strong (left) end of the scale. Fourteen programs (40%) received an average evaluation of "Strong" or better and 23 (65.7%) received an evaluation of "Strong" from at least one sub-panel. An additional seven programs (20%) received an average evaluation of "Adequate," resulting in 85.7% of programs assessed as "Adequate" or better. All but one program received an evaluation of "Adequate" from at least one sub-panel.

The frequency distribution for the other four ratings are presented below. The mean evaluation for these ratings ranged from 3.0 (program outcomes) to 3.27 (student quality). The modal response for faculty quality was 3.0 ("Strong"), for student quality and program efficiency 3.5 (between "Strong" and "Adequate"), and for program outcomes 2.5 (between "Excellent" and "Strong").


GSS Doctoral Panel Review: Mean Faculty Quality Rating
Frequency
0
2
4
6
8
Mean =3.19
Std. Dev. =0.916
N =35
1.00
Distinguished
2.00
Excellent
3.00
Strong
4.00
Adequate
5.00
Marginal
6.00
Weak


GSS Doctoral Panel Review: Mean Student Quality Rating
Frequency
0
2
4
6
8
10
Mean =3.27
Std. Dev. =0.741
N =35
1.00
Distinguished
2.00
Excellent
3.00
Strong
4.00
Adequate
5.00
Marginal
6.00
Weak


GSS Doctoral Panel Review: Mean Program Efficiency Rating
Frequency
0
2
4
6
8
1.00
Distinguished
2.00
Excellent
3.00
Strong
4.00
Adequate
5.00
Marginal
6.00
Weak
Mean =3.26
Std. Dev. =0.894
N =35


GSS Doctoral Panel Review: Mean Program Outcomes Rating
Frequency
0
2
4
6
8
1.00
Distinguished
2.00
Excellent
3.00
Strong
4.00
Adequate
5.00
Marginal
6.00
Weak
Mean =3.00
Std. Dev. =0.931
N =35

Program Directors' Opportunity to Respond and Transmission of Information to the Provost

Program directors were provided with the Program Reviews prepared by the sub-panels evaluating their doctoral programs on or about December 1. They were invited to respond in writing regarding any aspects of the Program Reviews prior to transmission of the Program Reviews and accompanying information to the Provost. Their responses were transmitted to Dean of Graduate Studies Marjorie Pryse in mid-December, for inclusion among the materials to be provided to the Provost—*i.e.*, the program-specific IRPE and AA data, the Program Documents prepared by program directors, and the Program Reviews completed by the GSS Review Panel sub-panels.

Reflections on the GSS Review Process

On December 8, the GSS Review Panel met to discuss and offer feedback about the Graduate Student Support review process, including informal recommendations by individuals for subsequent doctoral program reviews, when and if these should be initiated. The Provost attended the meeting and participated in the discussion. Panel members addressed several issues, although the Review Panel did not attempt to reach or formalize consensus about matters discussed. The agenda for the Dec. 8 meeting identified several aspects of the review process, as presented below. Not all of those aspects were discussed in detail at the meeting, although several instructive comments were offered. An attempt has been made to summarize many of those comments in accordance with the structure anticipated by the agenda. Panelists also spoke to several major and structural issues beyond the direct purview of the GSS review process.

[From Dec. 8, 2009 GSS Review Panel Meeting Agenda]

The GSS Review Process: Reflections and Recommendations

a. The Dimensions and Their Evaluation
   (1) Faculty Quality; Student Quality; Program Efficiency/Effectiveness; Program Outcomes; Overall
   (2) Evaluative criteria
   (3) The 6-point scale

b. Data Used in Evaluations
   (1) Institutional Research data
   (2) Academic Analytics data
   (3) Chairs'/Program Directors' Program Documents

c. The Evaluation Process
   (1) Two subgroups
   (2) Nine, three-member subpanels
   (3) Two subpanels per program, working independently
   (4) Program evaluations (subpanel reports): process, content, form
   (5) Comparing and synthesizing subpanel reports
   (6) Transmission to program directors/department chairs
   (7) Program directors' opportunity to review and respond
   (8) Transmission to Provost

A. ***The Programmatic Dimensions Evaluated and the Criteria and Rating Scales Employed in Their Evaluation***

- Information was provided by IRPE about student FTE generated by departmental faculty, including the percentage and total FTE associated with undergraduate students and the percentage of total student FTE generated by instructors who are graduate students. A question was raised about the relevance of this information to the GSS sub-panels' ratings of programs. It was suggested that the information about undergraduate FTE is potentially important to help explain departmental strengths or weaknesses on various dimensions of interest, including faculty productivity and the efficiency with which doctoral students progress toward their degrees.
- It was observed that the programmatic dimensions evaluated and the criteria used in their evaluation assume potentially critical importance in helping to establish the rules and benchmarks by which departments may be assessed in the future. The process not only is evaluative, but potentially prescriptive. It is important that the rules by which departments will be evaluated are established and publicized.
- It was suggested that further attention should be devoted to identifying precisely what is important in the evaluation and review process, and then developing templates that would guide both the provision and the assessment of information.

B. ***The Data Used in Connection With the Evaluations***

- Several panelists criticized the data and Faculty Scholarly Productivity ratings supplied by Academic Analytics. Widespread problems were identified with the AA data, including using inappropriate reference groups within which departments were ranked and comparatively assessed; reliance on inappropriate and/or incomplete indicators of faculty scholarship, including scholarly publications and research grants; using obscure or inappropriate criteria for identifying faculty within "programs" (which in many cases differed significantly from the departments assessed); and others. Some panelists observed that several program documents focused in large part on refuting or rebutting the accuracy and validity of AA data which cast their departments in negative light. Some panelists were of the view that the University should sever ties with AA and refrain from relying on the AA data because of their deficiencies. Others suggested that AA data were not meant to be employed for making inter-departmental comparisons, but rather were meant to be used exclusively for internal departmental reviews.
- It was noted that IRPE data did not in all instances match data kept by the programs, and that reliability checks are needed if IRPE data are to be used as measures of program quality in the future.
- It was suggested that Program Documents should include a section whereby program directors would identify specifically how they would invest additional resources, if provided, and how they would anticipate that such investments would benefit their programs.

C. ***The Evaluation Process Employed by the GSS Review Panel***

- Panelists expressed hope that the program reviews completed by sub-panels and forwarded to program directors would help those programs identify and respond to strengths and weaknesses, much as the external reviews completed in other

assessment processes. Some panelists expressed an interest in learning about the program directors' reactions to the program reviews, including their agreement or disagreement with points made in the reviews and more generally whether the reviews were perceived as useful or instructive.

- Some panelists expressed hope that the sub-panels' program reviews would identify opportunities for investment and measures that could be taken to strengthen programs (*e.g.*, "If this department had 'X' it could shine").
- It was noted that program directors did not receive all the data necessary for their preparation of program reports until the end of the summer, or a couple of weeks before their reports were to be submitted.
- It was suggested that program directors might require more time to draft program documents than was provided, especially if reliance continues to be placed on the often-flawed AA data, to enable them to put the AA and other data in proper context.

D. *Major and Structural Issues Beyond the Direct Purview of the GSS Review Process*

- Questions were raised about UAlbany's perceived position of disadvantage with respect to resource allocation compared to the other SUNY research centers. It was observed that a commitment of resources is necessary to cultivate graduate programs of excellence. Indeed, several panel members emphasized the importance of University and SUNY support to promote our doctoral programs and help them achieve or maintain excellence.
- It was suggested that several UAlbany graduate programs might be disadvantaged because they are not sufficiently large to comprise a critical mass necessary for attracting the best faculty, students, research grants, etc.
- It was suggested that difficult decisions must be made, including those in the nature of "scrap and build"; otherwise the entire University and all of its programs are apt to suffer.
- Different views were expressed about the strategy of "building on excellence," as opposed to "building on fundamentals" (placing greater reliance on "across the board" investments that may take time to flourish).
- It was suggested that "building on strength" is a less important objective than investing where the most value will be realized, including what return will be realized on investments.
- A need was expressed for the prioritization of University goals: *e.g.*, to serve undergraduate students; to achieve Tier II or Tier I university status; to promote excellence in graduate education.
- It was suggested that the infusion of additional resources is necessary and that other reform initiatives can be little more than "band-aids" absent those resources.
- It was suggested that several considerations are relevant regarding the contents and distribution of the GSS Review Panel's final report, including the understanding of the Panel members as they accepted their charge and assembled their sub-panel program reviews; appropriate consultation, information sharing, input, and decision-making authority involving University governance and University administration; and the potential effects on academic programs and the University. It generally was

suggested that the final report should identify the composition and charge of the GSS Review Panel; the procedures followed by the Review Panel; reliability measures regarding the sub-panels' ratings and a general description, including tabular and frequency distribution depiction, of the dispersion of ratings; and a section comprising GSS Review Panel members' reflections about the GSS review process.

- It was noted that another reason counseling against wider distribution of program reviews and program reports involves reservations about the validity of the AA data and the related concern that distributing the reviews/reports without also releasing those data invites the risk of misinterpretation or misunderstanding of the documents.
- Some GSS Panel members were agreeable to having all materials made available to the GAC, subject to appropriate confidentiality safeguards.
- The panel understood that the Provost would receive all supplementary materials (the GSS sub-panel reviews and the department chairs' responses to those reviews, access to the IRPE and AA data, and the departmental Program Documents), and that the deans of each School and College would also receive both the GSS Final Report, copies of the GSS sub-panel reviews for the programs in their School and College, and copies of the departmental chairs' responses to those reviews (and program documents to be shared with the deans directly by the departments). The Panel also understood that since they were created as a Provost's panel, it would be up to the Provost to determine further distribution of GSS panel materials, although the Panel understood that various Senate Councils would receive a copy of the GSS Final Report.
- Not all panel members expressed their views about the issues noted above and, as previously indicated, no attempt was made to achieve or measure consensus about the issues noted.