Performance Evaluations plus Optional GPA

DRAFT Proposal to stimulate DISCUSSION (1/13/2000)-- organized by Barbara Rogoff, with input from colleagues

Our deliberations about our assessment system amount to re-examining UCSC's institutional culture and educational philosophy. We need to think about what is best for students' learning, motivation, and preparation for careers, rather than either sticking with the status quo or slipping back into a conventional system that many educators agree has serious problems. Beyond the feedback we give students within our classes, the report that we give the 'outside' world tells the students what counts as the bottom line regarding their performance, impacting their attitude toward learning.

Institutional assessment systems have great weight in determining what is learned and taught; many national efforts to improve education focus on the need to revise assessment systems to support student learning (National Academy of Sciences, 1999; Resnick & Resnick, 1992). Because the weight of traditional systems (impinging from outside and from faculty and students' prior experience) tends to force efforts to improve education — like the NES — back into the mold, prioritizing a conventional grading system would likely destroy a narrative system. We need an assessment system of which we can be proud — one that contributes to our "uncommon commitment to education."

The question of revising our assessment system arises in the face of very real needs to reduce faculty workload. UCSC was cut to the bone in the early 1990s in such areas as staff support for faculty; while we are still recovering from that, we are faced with rapid growth that tests the limits of sustainability. The NES is not the root of the workload problems; although it needs revision, it has aspects that should be retained. The resource issues need to be examined more broadly.

This draft proposes, for discussion, that we

• **test our new optional GPA that allows students who need grades to have them** (and add pluses and minuses to that system, if there is general agreement), and

• **shift to a more efficient and simpler system of Performance Evaluations** which still facilitate multidimensional and contextualized assessment.

TESTING THE OPTIONAL GPA PORTION OF THE SYSTEM

A drastic move to institute mandatory grades does not seem warranted until we give the new optional GPA a chance to work. It has been in effect for only 2 years; no students have graduated under that system. Increase in use of the grade option by first and second year students compared with juniors and seniors (42% and 39%, versus 29 and 24%, according to Goff's report) suggests that this new option is beginning to have an effect. With time, it will also impact perceptions of the outside world, where many people are still incorrectly convinced that we "don't have grades." We should closely examine the optional GPA system in 3 years, when we will have data to determine whether there is a problem that needs solving.

There is not convincing evidence that mandatory grades would be an improvement. We could lose many outstanding students who choose UCSC <u>because</u> we provide multidimensional feedback without the necessity of grades. The <u>option</u> of grades probably attracts some students; <u>mandatory</u> grades would probably repel many. We would compete with numerous universities for students who want a mandatory GPA (rather than a choice); our competitive edge may be our unique niche of fostering academic excellence and creativity without relying so exclusively on grade-driven motivation. We attract many students who are interested in thinking outside the boundaries.

Narratives also seem to be a draw for many students to stay. Although some students who transfer mention dissatisfaction with tardy evaluations, their dissatisfaction with lateness does not imply that narratives themselves are a problem. And according to recent statistics from the Registrar, the problem with lateness is apparently well on the way to being solved.

We do need to inform colleagues elsewhere about our assessment system and how it serves educational excellence. The national press we have been getting provides such an opportunity, if we come up with a worthy system. In some disciplines, faculty feel embarrassed by colleagues elsewhere who criticize our system (and some propose that mandatory grades would solve this). However, many colleagues across the country who do research on learning admire UCSC's system that more closely serves the goal of enhancing students' learning than does the conventional system. Jerome Bruner, perhaps the world's foremost living scholar of cognitive development and learning, writes:

> "What IS this push to unidimensionalize everybody? I would have thought that Santa Cruz might have been proud of its narrative support system.... Simplify the narratives... but for God's sake, don't choose this sensitive moment in American history to give them up. Be proud of them!"

To convince the nation that we hold high standards for our students, it would be a step backward to regress to mandatory grades — a bureaucratic tool of the early 1900s — while assessment experts agree that student learning is best promoted by multidimensional evaluation in the service of improvement of understanding. (As Shavelson states: "a single grade inevitably collapses across dimensions to distort the full picture of performance [personal communication, 2000];" National Academy of Sciences, 1999; Shavelson & Ruiz-Primo, 1999. Problems with grades have been noted for some time — 41% of students at research universities nationwide reported that grade pressure makes it difficult to 'really learn something;' 43% reported that successful students 'beat the system' rather than studying [Boyer, 1987].)

It is not surprising for colleagues to question an unfamiliar system — this is the case for most innovations (Marris, 1986). Conservatism of this sort is a major reason why schools have difficulty acting on the findings on how to improve students' learning (National Academy of Sciences, 1999). Principles of learning are relatively well researched at this point — the challenge is helping educators understand them well enough to implement them. Indeed, the current uncertainty at UCSC likely stems in part from insufficient assistance to help newcomers understand the educational principles supported by narratives, and how to produce them efficiently. Most faculty have been hired in the last 7 years, and few have prior experience with narratives or help with them once they arrive.

To address colleagues' (and potential students') questions, we need to develop a clearer articulation of the rationale for our new system as it emerges. This could be provided by faculty who research learning and motivation, along with evidence from our students and alumni regarding the role of our assessment system in student learning, recruitment, retention, and career success.

There is no evidence that we need to make the radical change of forcing grades on students before we give the new optional GPA a try. However, an intermediate alternative we could consider is to give departments the authority to require their majors to take the grade option, if they are convinced that mandatory grades are an advantage to their majors.

SIMPLIFYING THE NES TO PERFORMANCE EVALUATIONS

Moving to Performance Evaluations would simplify the process of writing and reporting evaluations, through a focus on student performance and use of up-to-date technologies for streamlining the process. As discussed below, faculty workload would be reduced and the cost of the transition would not be large, with staff assistance to assist in implementation.

Value to student learning. There is little question of the educational value of an assessment system that

provides more information than a grade. The research indicates that evaluation systems that are more closely tied to the process of learning — focusing on the nature and quality of student accomplishments and gains in understanding — are more effective for fostering learning and motivation (National Academy of Sciences, 1999). Student motivation and understanding are promoted by environments that prioritize learning goals, in which students are encouraged to increase their competence and to figure out something new. This contrasts with environments that focus on validation of competence, where the goals become obtaining favorable judgments or avoiding unfavorable judgments of one's ability. Viewing a situation as an opportunity for learning (as opposed to an ability judgment) induces students to choose more challenging tasks, to perservere in the face of difficulties, to perceive instructors and peers as resources rather than as potential obstacles, and to foster and maintain longterm interest (Dweck, 1989).

Students and graduates are very articulate in describing the benefits of narrative evaluations for their learning and motivation. Their reports are consistent with the research literature:

•Narratives help them focus on learning the material to be covered (rather than simply playing a game of earning points). Motivation is fostered by an evaluation system that does not have a low ceiling (like inflated grades now do); narratives push students to go beyond minimal or even praiseworthy levels of evaluation to which grades are limited.

•Multidimensional evaluation aids students in distinguishing their strong and weak areas, and focuses their efforts on improvement in the quality of their work. For students whose performance in college is uneven — with strengths in some aspects of a course and weaknesses in others — a multidimensional summary of their work helps them diagnose what they need to be doing to improve, while giving them encouragement to build on their areas of strength. For many students from underrepresented groups or who are the first in their family to attend college, this feature of narratives is especially important. (The narratives also help them explain to their families what college is about, improving family support for their continuing in college.)

•A narrative system encourages students to learn from each other, rather than to treat each other as obstacles or to undermine each other. (Learning from peers is an important resource, especially as class sizes increase and make faculty less accessible. In addition, learning to collaborate is an increasingly important skill in the workplace [National Academy of Sciences, 1999])

•Across their college careers, students shape their trajectory using the narratives for information on areas of strength and weakness. The narratives at the end of each course provide information that they use in focusing on improvement for their subsequent courses. Instructors may regard their course evaluation as the end of a process, but for students the evaluations provide a distillation of a quarter's work that helps them improve across the years of college.

•Studying within a narrative evaluation system gives graduates excellent preparation for the kind of performance expected in graduate school and in jobs. Many have indicated that our system gives them an edge over students coming from conventional evaluation systems.

Value to instructional practice. The value of narratives for improving instruction is also noted by many faculty. In addition to being able to spend time talking with students about the course material (rather than explaining a weighting system — frequent in universities that prioritize grades), many faculty find that multidimensional narrative evaluations help them transform their instructional approach. However, with the large influx of new faculty, there have been insufficient institutional supports for helping newcomers understand the value of narratives for instruction.

The NES was key in improving my own instruction, especially in my large class of 140. After teaching for 15 years at a large university that offered only grades, I found at UCSC that the NES supported my change from

simply delivering information to the students to communicating with them and assisting them in developing their understanding and ideas. I had wanted to be a better instructor in my grade-based university; the NES provided a key tool for revising my educational approach by helping me focus on the multidimensional nature of the students' learning.

Bill Ladusaw recounts a similar experience:

> "What is at stake is not just the effect of the NES on students. There is also (or should be) a corresponding effect on how professors give feedback to their students. This is something that I experienced (with blessed relief) when I joined the faculty 15 years ago having taught at four other institutions under a grades only system. At that time, new faculty were well-instructed, at least in my department, on the nature and intent of evaluations. They were explicitly said to be NOT like letters of recommendation. They address only aspects of the student's performance in the class, not judgements of the individual. (E.g., 'X's performance in the course was excellent,' not 'X is an excellent student.')
>
> "It was due to the NES that I was able to reconstrue my job as a teacher -- to be a coach more than a judge, rewarding effort and assessing progress rather than (just) judging the level of attainment on some fixed body of material. It was the NES, I think, which made it possible (and humane) to teach beginning students to *do* linguistics rather than teaching them *about* linguistics. Free to challenge students to more sophisticated understanding about the methods with open-ended problems, I have been richly rewarded by some stunning performances (among the OK ones) that I don't think would have emerged under my earlier methods."

Moving to a focus on performance and simpler recording. The primary change needed — simplification of the process of writing and reporting evaluations — would be facilitated in part by recasting our focus from narrative evaluations to performance evaluations of the quality of student work. Some faculty regard a good NE to be like a long, personal letter of recommendation; many faculty who use templates or provide short NEs feel guilty or worry that others may be critical. We should revise the norm to respect template-based brief evaluation that contextualizes the dimensions of students' performance. Succinct and systematic evaluations would be honored along with flowing narratives. The aim is to move toward a norm where what is do-able is respected.

In a performance evaluation, the instructor summarizes the quality of the student's performance in the course as succinctly and systematically as warranted by the structure of the course and the instructor's goals for students' learning. This is very different than a letter of recommendation, where the writer is trying to evaluate the student and their chances of success in a different setting. In a performance evaluation, the writer simply summarizes the quality of the student's performance during the quarter, creating a report of the excellence (or lack thereof) of their accomplishments in the course.

Performance Evaluations resemble Performance Reviews that prevail in the world of work. (Consider why grades aren't used in the workplace.) In the workplace, performance is evaluated within the context of a job description, with an aim of fostering improvement. As Ladusaw points out, the fact that employees are told from the beginning which areas of their performance are to be reviewed forms the link between performance evaluations and improvement of pedagogy — what is to be evaluated at the end of the term must be clear at the beginning, in order to set up a sound recordkeeping system that focuses on aspects of the students' work relevant for evaluation. Using templates is a natural part of this process; a number of programs support construction of performance reviews.

Performance Evaluations could have a variety of forms, tailored to the instructor's goals. Instructors of very large courses whose courses are structured around a single type of performance might have one- or two-sentence evaluations (which may or may not refer to letter grades). For the largest classes, an instructor could provide a boilerplate description of the class, and a simple statement of the individual's performance, which could be entered automatically by the registrar's office, like this:

> The course was based on Sanchez's "Introduction to Chemistry" and covered basic concepts of inorganic chemistry and qualitative analysis. The grade is based on a numerical average of two midterms and a final exam. There were 250 students in the class, of which 60% earned a grade of B or better.

> Kevin earned a B-.

For those students whose performance stood out, comments would be added (e.g., "Sandra's contributions to discussion section were brilliant"). Although this is bare-bones, it is much more informative and adequate than a grade alone, as it indicates what the student's grade was based on.

Instructors whose courses are structured around several kinds of performance could provide a sentence or two on each type of performance. For example,

> Students in this semantics class were assessed on their ability to diagnose entailment, implicature, and presuppositional relations as well as their ability to apply analytical techniques to the analysis of lexical and sentence meaning through predicate-argument and quantificational analysis.
>
> > ---Tina's work for the course was mixed, with some excellent, engaged homework assignments and a midterm that showed very fine understanding of the material in the first part of the course. The grade of C reflects the fact that her work declined in quality sharply at the end of the course, with her final exam barely passing.
> >
> > or: ---John struggled with the formal aspects of assignments, though his homework and class discussion frequently revealed sophisticated insights and very fine ability to formulate relevant examples of linguistic phenomena.

For some classes, instructors could create free-form evaluations, but for many classes, a template would help instructors move easily and quickly from their records of student performance to a summary at the end of the quarter. In some departments, template models could be developed that faculty could adopt if desired. Most Performance Evaluations (see examples in Appendix) could be entered rather automatically using a template. Programs exist that allow course records to be transformed to written format automatically, according to the instructor's design.

We can encourage excellent use of Performance Evaluations as an instructional tool by recognizing instructors who go beyond the norm to develop evaluations that are exemplary in enhancement of student learning. An "Excellence in Evaluation" award could parallel our teaching awards, to focus on faculty use of Performance Evaluations to support student learning. Such an award would encourage ongoing faculty discussions that are necessary for the basis of an innovative system to be understood.

With such changes, Performance Evaluations would help us move from what is now sometimes seen as a punitive system to one that values and rewards excellence. (We also need to ensure that oversight of lateness is implemented fairly across campus.)

Student response to template-based evaluations has been very positive (according to a controlled study cited by Goff). Students appreciate the multidimensional evaluation of their accomplishments.

In my 140-student course, I hand out the template the second day of class to communicate the basis of evaluation; students are pleased that their accomplishments are reported systematically, and that they know the basis of evaluation at the beginning of the quarter. I avoid using grade-equivalent codewords such as "excellent," "good," and so on, and instead use phrases throughout the quarter that tie the student's performance to a standard and indicate the basis of evaluation (e.g., "Jan's essays were usually very well developed, with clear connections between ideas and evidence to support the arguments"). At the end of the quarter, the template yields narratives like this:

> Interdisciplinary course on the cultural basis of human development; involved extensive writing on a twice-weekly basis. 138 students were enrolled.
>
> > Overall, Afesa's participation and written assignments indicated a well developed understanding of the ideas of the course. We observed impressive progress in Afesa's understanding of the material, as evidenced by improvements in coherence and organization of ideas expressed. Afesa's 7 required 2-4 page essays and lab reports were usually of good, sound quality, reflecting active engagement with the topic, though in places the work would have benefitted from being pushed further; a few were very well developed, with clear connections between ideas and evidence to support

the arguments. The 6 required Reading Responses showed thoughtful engagement with the ideas. Afesa's presentation on Gender Roles showed thoughtful engagement with the ideas. Afesa went beyond the assigned work in the class, with an optional rewrite.

I give grades at the end of the quarter to the students who want them, but during the quarter, I make no reference to grades, so that students don't slip into a grade-based mentality. I hope to be able to preserve a learning-based 'climate' that way, even if a number of students request grades (this would probably be destroyed with mandatory grades).

The shift would be facilitated by a few staff consultants (at least for the first few years when the Performance Evaluation system is new). The consultants would be available to all interested faculty, to help them develop templates for their courses and input student records into their templates at the end of the quarter. They would help faculty use computer-based technologies to keep records during the quarter and to simplify the process of producing evaluations. (Some faculty have developed such systems themselves; others need staff support.) The consultants could also help orient new faculty to the system.

Transcripts should have a more professional format, instead of the current bulky, clunky ones. The Performance Evaluation transcript could be formatted to be concise and easy to read, with a professional look that will impress (and not overwhelm) readers. (Evaluations should also be accessible to faculty to aid us in advising and in considering applicants for research assistant positions.)

It would not be expensive to implement a Performance Evaluation system. The infrastructure of the current system is not costly and most of it would be needed for grades anyway, according to UCSC budget analysts and AVC Goff. In simplifying efforts (compared with the NES), Performance Evaluations would balance some of the cost of the consultants and awards. (In addition, compared with a grades-only system, there would be savings in the amount of faculty and staff time for writing letters of recommendation, since records for each student would be more complete.) If we create an academic assessment system that we are proud of and want to devote resources to, EVC Simpson has indicated that the faculty can indicate that this is a priority and resources could support it. In addition, a simplified and exemplary assessment system for the 21st century could be a rallying point for fundraising to support undergraduate education, helping us support excellence in education.

References

Boyer, E. L. (1987). *College: The undergraduate experience in America. Carnegie Foundation for Advancement of Teaching*. NY: Harper & Row, p. 144.

Dweck, C. S. (1989). Motivation. In A. Lesgold & R. Glaser (Eds.), *Psychology of education* (pp. 87-136). Hillsdale, NJ: Erlbaum.

Marris, P. (1986). *Loss and change*. London: Routledge and Kegan Paul.

National Academy of Sciences, National Research Council report of the Committee on Developments in the Science of Learning (1999). *How people learn: Brain, mind, experience, and school.* Washington DC: National Academy Press.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer.

Shavelson, R.J., & Ruiz-Primo, M.A. (1999). Leistungsbewertung im naturwissenschaftlichen Unterricht (Evaluation in natural science instruction). *Unterrichtswissenschaft, 27*, 102-127.

Appendix: A Range of Acceptable Performance Evaluations *(Names are pseudonyms.)*

Students were assessed on their final language proficiency by an ACTFL proficiency exam. The grade reflects the extent to which their skills improved during the course. This student's performance was as follows:

Listening comprehension: satisfactory

Speaking: poor

Reading comprehension: excellent

Writing: very good

Attendance and participation: excellent

Last class in three-quarter intensive introduction to organic chemistry. Text: *Introduction to Organic Chemistry* by Streitwieser/Heathcock/Kosower. Grading based on two midterms, weekly graded homework problems, and comprehensive final exam. Class performed extremely well overall: on final used 8 years ago, this class scored an average ten points higher than previous group.

Leon had strong homework performance. His final exam was at the class average while his first and second exam were above the average. A very good performance overall. Leon showed an unusual commitment to his education by enrolling in the Honors Program discussion section, in which students work in small groups on problems at and above those assigned as homework. Leon was an asset to the section by virtue of his thoughtful and insightful contributions to the collaborative process. His weekly collected and graded homework showed an excellent understanding of the material.

Course covers sculpting techniques in ceramics, issues and concepts related to the body and its representation, approaches to the artistic process, the relationship of contemporary art to studio practice, and professional practices.

Shawn's sculpting skills in this course were very good, his glazing and conceptual skills were good. His fish sculpture showed very good use of texture and visually interesting form. A suggestion is that he take more risks with sculpture by working on more challenging forms and concepts. Shawn was active in class discussions and participated in firing the kiln.

An upper-division lecture course in classical and molecular genetics based on the text by Griffiths et al. Evaluation is based on two in-class midterm exams and a comprehensive final.

First midterm: unsatisfactory, below passing level.

Second midterm: satisfactory, below the class mean.

Final exam: very good, among the top third of the class.

This overview neuroscience course dealt with: 1) molecular and cellular; 2) systems (sensory, motor, integration); and 3) behavioral (higher brain functions). Text: Kandel and Schwartz, *Principles of Neural Sciences*. Students evaluated mainly on test performances (three noncumulative midterms).

First midterm (100 points possible, 73 class average); Lisa's score: 81.

Second midterm (100 points possible, 71 class average); Lisa's score: 79.

Third midterm (100 points possible; 76 class average); Lisa's score: 74.

Approximate percentile: 69 (330 points possible; 244 class average); Lisa's cumulative score: 264.

Comments: Lisa performed quite well on the unannounced quizzes. Based on her quiz scores it is evident that she showed very good preparation for the course lectures. Overall, Lisa exhibited sound mastery of the principles and facts of the neurosciences in this relatively demanding course.

In this interdisciplinary composition and rhetoric course, students explored intellectual and imaginative tensions between language and environment. Six 1000-1500 word analytic essays covered a range of topics.

Tu Anh's performance was remarkable. She participated intelligently in discussions, submitted every reading log assignment, and demonstrated an exceptional grasp of the material. She also engaged in significant extra-credit reading and proved to be a respectful and astute editor of other students' work. Her formal compositions were strong and interesting: analytically keen and conceptually ambitious. Her first essay, for example, effectively layered poignant anecdote with relevant textual references in exploring the meaning of a group camping experience. Her fourth essay impressively critiqued the ways that Jamaica Kincaid and Walker Percy believe that humans have lost contact with pure ways of living in the world; her inclusions of Taoist theories of the self were deftly and aptly integrated. Her excellent skills in argument were exhibited in her fifth essay -- a consideration of animals and laboratory research that was particularly striking in its balanced and in-depth examination of science and spirituality as cultural constructs; her subsequent revision seamlessly incorporated a brilliant reflection on Vizenor's Chippewa novel, *Dead Voices*. In short, an outstanding intellectual performance for the quarter.