

pline, proceeds largely in isolation from preceding work by focusing entirely on instruments of analysis from another discipline and in disregard of work done by political scientists. Given such discontinuities I find it difficult to designate as "institutionalization" the development of comparative politics over the last forty years.

The difficulties comparativists confront are much like those of the wider discipline. As David Ricci (1984) has suggested, we are engaged in two incompatible enterprises: the study of public life in a scientific fashion, and a devotion to a particular set of political norms. But we also have careers to pursue. Are we trying to serve too many masters?

---

## Note

\*Prepared for Mini Plenary Session VIII.2, "Roundtable on the Institutionalization of Comparative Research," International Political Science Association, 24th World Congress, Washington, D.C., 1988.

---

## References

Almond, Gabriel. 1983. "Corporatism, Pluralism, and Professional Memory."

# Between Leibniz and Voltaire: Exams and Grading in a Less Than Perfect World\*

Donald Chisholm, *University of California, Los Angeles*

Leibniz contended that out of all the possible worlds that God could have created, God elected to create the best one. In my best of all possible teaching worlds I would be working with a small group of students, all highly motivated, with the basic intellectual skills necessary for success; the reading list would be extensive but well chosen, the class would meet weekly, with the students avidly vying with one another to express their thoughts on the subjects at hand. In fact, they would be as interested in the material as I am. When each class finished, students would

- World Politics* XXXV: 245-260.
- Andrews, Williams G. 1987. "The Impact of the Context on Political Science in the United States: The Formative Years." Paper presented for Conference on the Comparative Study of the Development of Political Science, Cortona, Italy.
- Berndtson, Erkki. 1987. "The Development of Political Science: Methodological Problems of Comparative Research." Paper prepared for Conference on the Comparative Study of the Development of Political Science, Cortona, Italy.
- Diamant, Alfred. 1962-63. "The Field of Comparative Politics." Lecture delivered to regional seminars sponsored by the American Political Science Association, Syracuse, New York; Reno, Nevada; Bloomington, Indiana; and Knoxville, Tennessee.
- Gunnell, John. N.d. "The Historiography of American Political Science." Mimeo.
- Haeberle, Eckehard and Robert Picht. 1980. *Situation of Comparative Political and Social Research in Countries of the European Community*. Brussels: European Cooperation Fund.
- Ricci, David M. 1984. *The Tragedy of Political Science: Politics, Scholarship and Democracy*. New Haven: Yale University Press.

---

## About the Author

Alfred Diamant is a Professor at Indiana University. His areas of expertise are Comparative Western European Politics and Social Policy.

still be enthusiastically discussing the week's material. There would be no formal examination and no grades. To provide pedagogic structure for their efforts and to help them focus their thinking, the students would be asked to write papers weekly, the precise topics of which would be selected in consultation with me, and those papers would go through several drafts, until they were highly polished, shining pieces of work. In order that I might oversee the development of my students, they would take courses with me for at least a year. At the same time, my course

load would be light enough to permit the investment of time and effort necessary to teach such courses and I would be rewarded professionally (i.e., promotion and salary) in a way that would encourage me to make that investment.

Voltaire disagreed with Leibniz's assertion. And I cast my lot with Voltaire on this issue, at least as it applies to examining and grading students. Most of us will recognize that not even for high-quality liberal arts colleges, let alone the larger public universities, does the above description have even the most remote applicability. Classes are not small. When we teach large courses, as often we must, students' work is evaluated by graders or teaching assistants—distancing us from our students. It is the rare exception when students take more than one course with us. We usually are able to follow their development, if at all, only indirectly. Students are often not highly motivated. Although subject matter and the reputation of the instructor enter into the calculus at levels significantly above zero, the decision to take a course often hinges on when it is offered, the size of the reading list, how many times weekly the course meets, and whether it fulfills a requirement. It is the exceptional not the model student who reads through all of the course material. And most of us are unfortunately familiar with the "iron law of oligarchy" that governs student participation in the classroom. Worse, many students have only rudimentary skills, inadequate to the demands of college coursework. We rarely see students at office hours unless they have problems of some sort. And, whereas in my best of all possible worlds, course assignments serve only the positive purpose of focusing the learning experience for the students, in the real world in which I teach, that is but one function that must compete with several others. Grades *do* matter and must be assigned. I never tell students that grades are unimportant—it is dishonest to do so. Course assignments serve also as coercive mechanisms to encourage students to do more than dabble in the reading. Additionally, exams act as heuristic devices, telling the instructor if what one thought

was being taught actually was being received and understood by the students.

Moreover, professional incentives within the university setting do not weigh teaching heavily. Lip service may be paid by university administrators and colleagues to the importance of teaching; in most cases, however, research and publication remain the *sine qua non* of advancement and salary. The problem is that structured time—teaching—drives out unstructured time—research and writing. That we recognize this basic disjuncture is evidenced by the sustained effort to reduce teaching loads and the admiration typically expressed for those academics who achieve this goal. To be sure, many of our colleagues expend great energy on teaching, some from a deep sense of obligation, some out of the rewards intrinsic to teaching; few do so because of the external incentives offered in academic institutions.

The combined effects of the actual environment in which we find ourselves teaching, if not as grim as the world that confronted *Candide*, constrain our choices about the kinds of course assignments we make, the way we evaluate them, and the manner in which we deal with students regarding them. These constraints lead to predictable patterns of behavior on the part of instructors, which can be counterproductive to good teaching. The characteristics of the environment are unlikely to change any time soon; the underlying causes will remain. How then to ameliorate the effects of those pathologies? I suggest here some tactics to help attain satisfactory results in the face of these considerable structural constraints.

### Designing Exams

Let me lay out my exam biases straightaway. If a liberal arts education is to be worth anything at all, students must learn to write, for clear writing is clear thinking. Deciding what to write and how to write it is no less than problem solving. All of my courses, whether introductory or advanced, include essay examinations. Most include papers as well. They may or may not include short-

answer exam questions; only once have I used multiple-choice questions, and then in concert with essay questions. More on that shortly.

The initial problem revolves around how many exams to give and at what points in the term. My experience has almost exclusively been with the quarter system; for semesters I expect the answers would be somewhat different. With the brevity of quarters, I generally give one mid-term along with a final exam. The key for the mid-term is to give it late enough to permit the students time to master a significant segment of the material, but early enough to have sufficient time to modify their study habits (if warranted) before the final. It is only fair to make clear to the

---

*The problem is that structured time—teaching—drives out unstructured time—research and writing.*

---

students at the outset the number, timing, and character of the exams—and to stick to that regimen.

Composing questions for examinations is most usefully approached in a manner analogous to designing items for use on surveys. Criteria demarcating good from bad questions are much the same for both applications. A good exam question should test students' comprehension of some key aspect of the course, arraying them across the entire spectrum of possible grades, presumably approaching the distribution of a normal curve. If the question is too difficult, no one does well; if too easy, the converse obtains. While I do not grade on a curve, I read a sample of exams before beginning to grade to determine if the questions I asked were reasonable. Although I may have an ideal answer in mind, I temper that vision with the reality of what the students were able to accomplish, adjusting my grading standards accordingly, up or down. In a similar manner, questions may not spread students across the possible range of scores because the questions

were poorly designed and did not effectively test mastery of the key components of the course. Such questions are not useful because they do not discriminate among good and bad students.

In courses in which a paper is not assigned, I often mix a take-home exam with an in-class exam so as not to disadvantage students who are skilled at papers and less proficient at in-class exams. Take-home exams are fraught with their own problems. To minimize them, I let the students pick up the exam on either of two days, any time during working hours, with the exam due back precisely 24 hours later. I give specific instructions that the students are to write no more than a set number of hours, and I enforce a strict page limit for their answers. This approach provides reasonable flexibility for students who are employed or who study best at odd hours. At the same time, "red hots" are prevented from surpassing other students simply by dint of hours invested. The net results have been good quality answers and a fair measure of student satisfaction. Moreover, as I require them to be typed, they are easier to read than hand-written in-class exams.

In my undergraduate courses, I have, with few exceptions, employed a mix of essay and short-answer questions. The former assess the students' grasp of key course concepts. The latter serve the coercive function of ensuring that the students pay close attention to the readings and the lectures. I weight the essay questions about two-thirds to three-quarters, with the short answers consuming the rest, with the rationale that to pass the exam, a student must do more than regurgitate, and, to do really well on the exam, a student must know the details as well as the general outlines. Put differently, the essay questions reinforce for the students the material one hopes will "stick to their ribs." That is, it is possible to make exams into an integral part of the learning process. In this sense, it is not only unfair, it is never useful to throw students "curveballs" on exams.

It has been my experience (both as student and instructor) that few students are capable of writing much of

## The Profession

merit after two hours. This time constraint suggests that it is impossible to test students on all aspects of the course, and one must choose important/representative questions.

Moreover, each essay question is designed so that to answer it competently a student must have mastered key concepts in the course and be able to apply them to concrete situations. The best students will understand that to answer the question they must first deal with definitional and other ancillary issues not explicitly demanded by the question. They will then deal with the general concepts and demonstrate their mastery of them by providing relevant examples for illustration. Often times, I break the essay questions into several sub-questions in order to give cues to the students about how they might best approach an answer. I have also asked the same question in two very different ways, allowing students to answer the version that resonates best for them. In courses addressing theoretical issues, questions that begin with an assertion, then query the students about that statement, have proven useful. I generally prefer to ask "hard" questions that aim to provoke thought and then evaluate answers to them fairly leniently, rather than ask "easy" questions and grade like Attila the Hun, the scourge of God. Quite by accident I discovered that when I placed substantial demands upon the students, they responded by taking the course more seriously. The potential tradeoff lies in the propensity for students to equate effort expended with achievement; i.e., if they work hard, they deserve higher grades.

I have generally refrained from using "objective" exam questions—multiple choice and/or true-false—in favor of short-answer questions that require students to write, both identifying the term, concept, or person, *and* explaining its significance. The obvious trade-off is that one can ask far fewer such questions than if they were multiple-choice format, and accordingly they must be more carefully representative of the course content. Where once I gave students relatively little choice, I now permit them fair leeway, say six of ten or twelve terms: their grades are usually no better but it gives them more control over their own fate.

Over the years I have become persuaded of the value of distributing review questions the week prior to both mid-term and final exams. It genuinely seems to help students focus their study. Additionally, it equalizes the exam intelligence available to those students outside those social organizations keeping exam files. I typically will hand out a week prior to the exam a review sheet comprised of about 20 essay questions, that number in order to reduce the temptation to do no more than write out answers to all the questions as the sole way of preparing. And none of the review questions shows up on the exam.

Finally, given the nature of the subjects I most often teach, I find those final exams are most fair and useful that cumulate material from the whole of the course. That is, my finals overlap with the mid-term exams but emphasize material covered in the interim. I tend to weight the mid-term and final about 40% and 60% respectively, on the assumptions that most students will not comprehend the material till well toward the end of the term and that they ought to be given the opportunity to improve on their mid-term grade.

## Related Problems

I operate from the simple assumption that universities and colleges have no particular monopoly on morality, that on every campus some will work outside the prevailing norms to what they think is their own advantage. Nor do I harbor any illusion that an exam system can be immune to all forms of dishonest exploitation (and there are many). At the same time, we have an obligation to our students to maintain at least a minimum standard of fairness. I therefore try to structure my system so that the cost of academic dishonesty is high enough to dissuade all but the most determined from pursuing that course.

Student social organizations often maintain files of previous exams given by particular professors. To obviate this problem, one might collect the questions after each exam, but I write an entirely new exam each time I give a course. When one

gives closed-book, closed-note exams in-class, crib notes can be a problem. One might permit a single 3 × 5 card for each student, on which may be written anything the student thinks may be helpful. Another tack is to provide examination blue books, imprinting each book with a unique stamp. One may also ask students to start on the second (or whatever) page, in order to avoid "cooked" blue books, an inevitable temptation if exam questions are given out in advance or if the same exams are given repeatedly over the years.

A related set of problems revolves around turning in take-home exams in or taking exams at the specified time. In most courses, some students will have exam scheduling conflicts. It does not seem fair to ask a student to take exams back to back, and so I most often will permit them to take an alternate form of the exam at some other time. It creates more work, but seems more equitable, and probably produces better exam performance than compelling students to take the exam only at the appointed time.

And there will always be students who cannot or will not turn their papers in on time. I have ceased trying to fathom which stories are true and which conjured up (Did their grandparent really die? Were their notes stolen out of their car?). Instead, I simply tell those students that the paper (or whatever) is due by the final, or they will fail the course. I confess to some satisfaction from stopping students in the middle of their laboriously devised, elaborate explanations by explaining that it is irrelevant and all I want is the paper. Whether one should penalize assignments turned in late is a separate issue. My rule of thumb is not to do so unless the student appears to be working overtime to manipulate me to his or her advantage.

## Grading: Problems and Meliorations

After designing the exam, I typically write in outline form what I construe as an excellent answer to each of the questions. This clarifies in my own mind what I expect the questions to address and provides a

template for evaluating student answers. However, I have found it wise not to adhere too rigidly to the template: students sometimes show me legitimate ways to answer the questions I had not considered and, after reading a sample of exams, my template may prove unreasonably difficult.

Most of us are familiar with the problem that might best be termed the “inevitable regression toward the B- /C+.” This phenomenon is a function of several factors, chiefly fatigue. The more exams one grades with rest, the more difficult it becomes to differentiate among them as to quality. Related to this is the common human desire not to disappoint others, in this case assigning students low grades. Conversely, we know that we cannot grade too leniently. The B- /C+ range offers a relatively comfortable compromise between these extremes. In order to lessen this tendency and its deleterious effects, several tactics, all addressed to the problem of fatigue, seem to work. I grade no more than five essays in a row, taking a decent break between groups of that size. Additionally, I tend to switch back and forth between the different questions answered by the students. The cost is that refocusing is necessary; the benefit is that one is less likely to become jaded in evaluating any one question.

In the best of all possible worlds, comments on exams serve an important pedagogic function. But comments take considerable time and effort if they are to have any meaning. One soon wishes for a thesaurus of comments to avoid repeating phrases. And, in the world we inhabit, writing comments on exams threatens to become an exercise in justifying the grade assigned. The worse the grade, the more comments are needed for its defense: when a student comes in to ask about a grade, we must be able to justify our action. Thus, a subtle shift occurs in the character of comments: they tend to move toward the negative side. A concomitant is that the faster one grades, the higher the grades assigned are likely to be; it takes more time to justify lower grades than higher ones. Even though one might have the deep conviction that an answer

deserves a lesser grade, assigning a better one is easier.

It requires constant vigilance to avoid these pitfalls, but it can be done. Comments are especially important on mid-term exams, because they tell the students whether they have been learning the material properly in time for them to make adjustments for the final exam. I try to put comments along the body of the answers and summary remarks at the end. By the same token, I never put comments on final exams, as most students do not pick them up and, for those that do, I speak with them in person about their performance. As with the regression to the B- /C+ range, fatigue exacerbates these tendencies. Allowing enough time between when the exam is taken and when one promises to return them so that one can grade not more than a certain proportion per day is useful. The discipline required to so apportion grading efforts is difficult to attain, but the improved quality of grading suggests its worth. The very real danger in spreading out grading over, say, a week is that one’s standards will subtly shift as the days pass, so that exams graded at the end will be graded differently than those at the beginning. My tendency is to grade the last exams more leniently than the first. This propensity may be counteracted, partially at least, by rereading exams initially assigned the several grades.

Fairness in grading is also an issue. Problems of grading students in an even-handed manner cut both directions. There are students for whom one roots and hopes will do well. There are also occasionally students, who for one reason or another, I simply cannot abide. In general, I try to grade “blind,” that is, without looking at the name of the student. But for students whom I find truly problematic, I ask a colleague to grade the exam for me, so that even the appearance—let alone the substance—of impropriety will be avoided.

In large lecture courses, one may have one or more readers or teaching assistants responsible for grading. Although this frees us from the burden of grading, it prevents direct communication from student to instructor, which composes a signifi-

cant part of student motivation—to know that their professors will read their thoughts. It also introduces the potential for wide variation in grading. In order to maintain some standardization, I write an outline of an excellent answer for each question, copies of which are distributed to the teaching assistants, and discuss each question with them as a group before they begin grading. I remind them to praise students who have done well and to find something good to say about all but the worst exams. I also encourage them to talk with each other about the grading to facilitate some reasonable adherence to the same norms. After the exams have been graded, I have sometimes asked each to give me some representative exams from each grade category, so that any gross disparities can be remedied before the exams are returned.

### After the Exams Are Returned

In almost every course some students will write failing exams, or exams that at best merit no more than a “D.” For such students, I do not assign a grade immediately. I request that the student come to my office hours to discuss the exam. I try to gain a better sense of what the problem was (lack of inability, inattention, personal matters), recommend some effective remedial steps in studying and writing that the student might take, then I ask the student to rewrite the exam and return it for grading. This approach seems to work fairly well; although dramatic improvement is infrequent, usually there is some change for the better, even if for no other reason than a “Hawthorne effect.”

In any given course a certain percentage of students will be unhappy with their grades. To be sure, some students will come to office hours to learn how they might improve; others will have genuine complaints, as errors in grading are virtually inevitable. But a significant number will come only to see if they can garner a few extra points. In a sense, we create conditions that encourage such behavior. When we assign points to essay questions, we create an illusion in the minds of our stu-

dents that we can grade to the precision of discriminating between different answers by a single point. This not infrequently results in histrionics, manipulation, mau-mauing, and general unpleasantness. I doubt seriously that anyone can grade an essay much more accurately than within half a grade. I have also discovered that most students end up with mid-term and final scores that fall within a range of one-half grade of each other, suggesting that my grading is reasonably accurate. Those falling outside that range I usually recheck to see if I have made an error.

Several tactics address these problems. One can simply assign letter grades only to essay exams. Upon returning the exams I always go over in some detail what would constitute an "A" answer to each question. I make it mandatory for all students requesting a reevaluation of their exam to submit a one-page written argument as to why they deserve that change. I reread their exam and respond in writing. If the student is still unhappy after my response, then I will meet personally with that individual.

If I am unwilling to change the grade and the student is adamant, then I have occasionally resorted to the cruel expedient of having the student read the disputed answer to me aloud. That's usually all it takes to end the dispute. I have also found it helpful to photocopy excellent essays, sans student name, to show to stu-

dents with queries about their grades. This is more effective than having them read my own template because it demonstrates what one of their colleagues was in fact able to achieve, not merely what I expected. Another approach is to enforce a 24-to 48-hour "cooling off" period before students may come in to discuss their exam results. On the other hand, given the relative imprecision attendant to grading essays, I am usually willing to grant a few more points.

For students who genuinely wish to improve their performance on the next exam, I do not insist on any of the above; I see them directly. If the course has graders or teaching assistants, students are required first to go through the process with them, before seeing me, in order to minimize the opportunity to play off instructor against assistant. Except for cases in which errors were made in computation, I do not change grades after grade reports have been turned into the registrar.

Following a practice I learned as an undergraduate, I ask those students who have done especially well on an exam to stop by my office hours to chat. This permits me to extend my felicitations on their performance, especially important for students in large, anonymous classes. It gives me the chance to see my best students, helping to balance those more frequent interactions with students who only have problems.

## A Few Last Words

I have touched on some of the difficult problems associated with exams and grading in this less than perfect world in which we teach. I do not pretend to have surveyed them exhaustively nor to have provided any definitive solutions. Indeed, I do not believe there are solutions, only palliatives that may reduce the violence of endemic problems. The suggestions made here result from my own trial and error (with emphasis on the latter) experience with undergraduate courses in several large public universities. Others will no doubt have additional, and perhaps more effective, suggestions, which fit more closely the contours of their own corporate cultures. I do not think I am overly pessimistic in asserting that these problems will never go away, nor excessively optimistic that we can find effective mechanisms for their redress.

---

## Note

\*My thanks for their helpful comments on an earlier version of this paper to Gregory Caldeira, Nelson Polsby, Richard Sklar, and John Zaller.

---

## About the Author

Donald Chisholm is a member of the Department of Political Science at the University of California, Los Angeles.

## Public Choice in Political Science: We Don't Teach It, But We Publish It\*

Jay Dow, *University of Texas at Austin*

Michael Munger, *University of North Carolina at Chapel Hill*

The study of political science has been substantially transformed by the work of public choice scholars over the past 40 years. Such works as Arrow (1951, 1963), Buchanan and Tullock (1962), Black (1958), Downs (1957), Niskanen (1971), Olson (1965), Riker (1962), Sen (1970), and Stigler (1971) now appear extensively in the references of articles published in a variety of our professional journals. We were interested in discover-

ing how extensively the public choice approach is being pursued in graduate political science programs in the United States. Our focus is mainly on programs in American politics, because this field more than any other has been the forum for work in public choice. To foreshadow our conclusions, the results tend to indicate that public choice in political science is seen as an offshoot, or related discipline, rather than as a

substantive field in and of itself. The authors of this paper feel this is a mistake, and hope to persuade the reader.

We offer two arguments for this position. First, as we demonstrate later, a significant proportion of publication in our discipline's professional journals take a public choice perspective. Students with no introduction to the jargon and methods of this approach are needlessly