Chemical Education Research

Teaching the Sophomore Organic Course without a Lecture. Are You Crazy?

Alexander Z. Bradley, Scott M. Ulrich, and Maitland Jones Jr.*
Department of Chemistry, Princeton University, Princeton, NJ 08544; *mjjr@princeton.edu

Stephanie M. Jones
Department of Psychology, Yale University, New Haven, CT 06520

There are several reports in the literature on efforts to use cooperative learning in general and physical chemistry (*1, 2*). Fewer attempts have been made in organic chemistry, although the Workshop Chemistry Project of the Rochester group is a notable exception (*3*). In this paper we provide a description of cooperative learning in the organic course in which the lecture component diminishes in the first semester and essentially vanishes in the second. It is surely legitimate to ask straightaway what we can possibly add to the already prolific, if not prolix discourse. We see two possibilities. First, our venue is the sophomore organic course (*4*), notoriously difficult and pressure-packed. Is this any place in which to experiment? Second, in year one of our cooperative course we were able to control our experiment in a way that we think has not been done before and which is therefore unusually informative. So we hope that this report will be more than an anecdotal description, will break some new ground, and will point the way for future improvement.

Background

When we began this experiment there was but one organic course at Princeton;[1] all students, regardless of their future intentions, took the same large, traditional lecture course. The class is about 80% sophomores, with a leavening of 20% freshmen who had achieved a grade of 5 on the College Board Advanced Placement (AP) examination. There is also a sprinkling of juniors and seniors. Traditionally, there were three 50-minute lectures a week; a fourth "hour" was a recitation section taught by a graduate student. Marc Loudon's fine text (*5*) was used until an in-house book (*6*) became available. We gave three "hour" examinations each semester, although we allowed 2.5 hours for their completion. There was a three-hour final examination, which included some choice. Questions were of the essay and problem-solving type; there were neither multiple-choice nor short-answer problems. The examinations were graded by the teachers in the course, the director of undergraduate laboratories, and the graduate student teaching assistants. Although Princeton is a highly selective university, the organic course seems to share the usual problems peculiar to the organic course everywhere: a certain reputation as being a killer, hyper-anxiety and over-competitiveness on the part of some of the students, and a stubborn refusal on the part of many students to believe that any attempt to deal with this course by memorization would ultimately be fatal.

In the fall of 1996, one of us taught a seminar to 16 super-qualified freshmen (AP5), who were presumably immune to any deficiencies in teachers or their teaching methods and were thus appropriate subjects for experimentation. One of the most successful of these experiments was the introduction of small-group sessions in which students worked on problems cooperatively. Naturally, the question arose as to whether group cooperative learning could be exported to the big course. This activity had been so successful in the seminar that it seemed worth the effort and risk to try.

Course Structure

The new course was taught by a longtime teacher of the traditional lecture course, a teaching/research postdoc, and a prize-winning graduate student teacher. We prepared the students by running the first-semester graduate-student-led recitation sections in the group problem-solving format. About half way through the first semester of the traditional course, we called for volunteers for an experimental 60-person second-semester course to be taught at 8 a.m. (!) entirely in the small-group cooperative format. Somewhat to our surprise, we were overwhelmed with applications: more than half the first semester pool of 230 applied. As in most volunteer groups, the applicants were disproportionately from the top of the class. Because we wanted to approximate the normal distribution in the course, we filled from the bottom of the grade list: if you did badly in the first semester, you got in; if you got an A, you entered a lottery. In the end, we still had to accept a class that was better than the class as a whole in terms of first-semester grades.

For comparison purposes, this "X" group was paired with a control section in the larger regular second-semester lecture course, which was taught in the traditional lecture style by one of us. The control group was selected so that they had received the same grades in the first semester of the course as had the volunteers in the X course ($t[124] = 0.73$, $p > .1$), and the members of neither the control group nor the regular section yet knew of the control group's existence. The two groups took *exactly* the same exams under *exactly* the same conditions. The members of the X section were asked several times not to attend the regular lectures. A few people ignored our pleas on occasion, but as far as we could tell, no one routinely "double dipped". The problems given in the X section, and their answers, were posted at the end of each week on the course Web site, so the students in the regular class had access to this material. During the semester there were 12 joint question sessions to which both classes were invited.

The experimental section was divided into permanent small groups of four or five. A fair amount of social engineering was applied so as to create groups that were balanced in terms of grades in the first semester and gender. One student in each group was designated as the scribe, whose responsibility was to write up the answers. The position of scribe rotated each meeting. The work of the semester centered on written problems handed out at the beginning of each class. Students were expected to have done the reading, which was assigned through the Web site. Over the course of the semester, we worked 62 problems ranging from the quite simple (few or no "drill" problems were given) to the truly brutal. We tried to have the problems evolve through the week, with more simple problems leading to more complex ones, and we tried to keep a balance between synthesis and mechanism problems throughout the semester. The groups would work on the assigned problem until we and they were satisfied with the answer, and then the daily scribe from the group would write the answer in a book. We collected the books and corrected and graded them. On several occasions, we rotated the books among the groups and had the students correct and grade their neighbor's answers. There were only three regular lectures,[2] generally on mornings after the evening hour exams. Minilectures of 10–15 minutes were sometimes used to open or close classes. Grades were assigned on the basis of three hourlong examinations (dropping the lowest grade) and the final. There was no grading of the group as a whole.

The three of us spent the 50 minutes working the room, wandering from group to group. We watched how the groups were working, listened to what was being said, and looked at what was being written. When a group was stuck, or moving relentlessly down some unproductive path, or had had some unusual idea, we would stop and talk to them, questioning, redirecting, encouraging, congratulating, and even admonishing from time to time. In practice, each of us talked to every group at least once on a given day. If a particularly interesting or difficult point came up, there might be many conversations between one or more of us and one group. A widespread difficulty among the groups would send one of us to the blackboard for a five-minute microlecture. The constant feedback from the students allowed us to adjust the problems, not only for future class meetings, but even in the middle of the class period. Ad hoc extensions or additions could be written on the blackboard, and we were able to bring up new points or to follow especially interesting ideas "on the fly".

Our roles were vastly different from the role of the traditional lecturer. A traditional lecturer attempts to show students a path through the material. If he or she is good at the job, a sense that there are many approaches to success in this rich, diverse discipline will come through; but there can be no escaping the difficulty of encouraging different approaches to learning the subject matter in a single lecture course. That is not to say that lectures have to be unidimensional or the same each time through: the proper model for the lecture is not classical music, but jazz. The accomplished lecturer improvises on the subject material. Each time through will be fairly similar in overall content, but vastly different in detail. You will recognize the song, but should be able to enjoy each new performance. In the problem-solving format we were in charge of 16 improvising groups, each of which was playing the tune in a different way. Clearly, this teaching style puts severe demands on the teachers, but there are enormous opportunities as well, precisely because of the ability to encourage several different approaches at the same time. We find ourselves solidly in agreement with Howard Gardner (*7*) in the sense that we don't see our job as transmitting a given factual canon to our students, but rather of helping them to develop the skills that will allow them to function successfully within the discipline. Of course, we have to speak to each other, and there is a grammar to chemistry that must be learned. But there is now a factual ocean out there that is beyond the ability of most of us to remember. Success requires being able to generalize and to see as similar things that initially appear quite different. There may be only one way to memorize, but there are many pathways to what we might call, with more than a slight risk of hubris, enlightenment in chemistry (*7*).

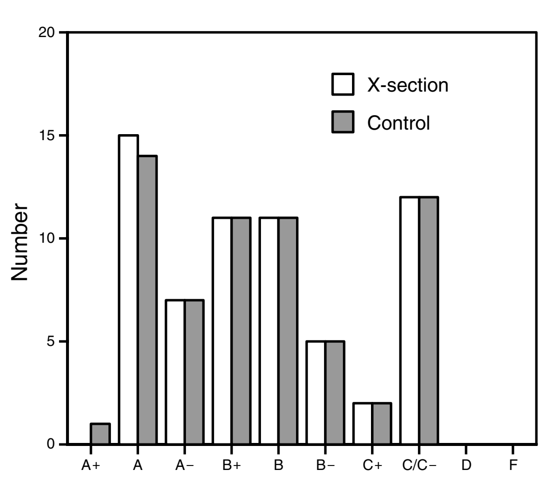

Figure 1. Grade distribution for the X section and the control group in the first-semester lecture course.

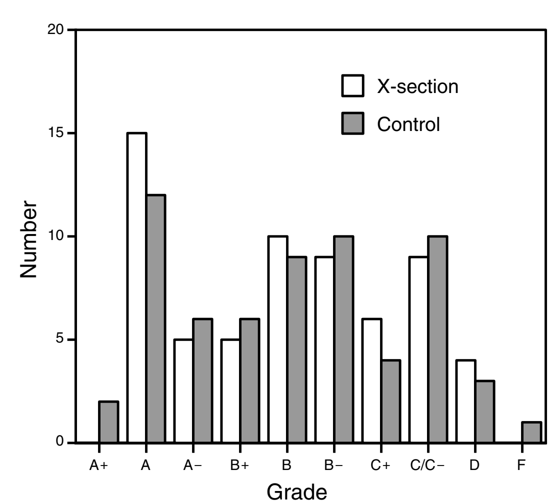

Figure 2. Grade distribution for the X section and the control group in the second-semester course. The X section did group work and the control group took the lecture course.

What Happened in the First Year

The take-home lesson can be couched in two ways. One might say, for example, “We *did* manage to teach this incredibly difficult subject without any lecturing!” Or one might equally well say, “It makes no difference whatsoever how you teach this material—some folks will get it, others won’t.” In fact, the X group and the control group performed exactly the same throughout the year. At no point was there a significant grade difference between them. Sometimes the X group was slightly better, sometimes the control group, but they were always the same within the noise level. For example, using a multivariate analysis of variance (MANOVA) with the three hour tests and the final exam as the dependent variables and control versus experimental group as the between-subjects factor, neither the omnibus F test ($F[4, 121] = 0.698$, $p > .1$) nor any of the follow-up univariate tests ($F[1, 124]$ ranged from 0.073 to 0.506 for the four dependent measures, $p > .1$) for group membership were significant. Moreover, a repeated measures analysis of variance revealed no difference between the two groups in change in grades across the semester ($F[3] = 0.91$, $p > .1$).

Figures 1 and 2 show the final grades for the X and control groups for the two semesters of the course. It would seem that the factors determining grades are largely independent of teaching method. In the first semester the mean grade in the X section was 71.96 (SD = 10.55) and in the control was 73.34 (SD = 10.77). In the second semester the X-section mean was 64.72 (SD = 12.72) and the control section mean was 65.01 (SD = 15.04).

In their assessment of the guided-inquiry method applied to general chemistry, Farrell, Moog, and Spencer came to a closely related conclusion because their experimental group performed roughly the same as the traditionally educated students in subsequent courses (*2*). Although the guided-inquiry method does not closely resemble our experiment, their results do reinforce the notion that it is not easy to (constructively) induce grade inflation!

Also revealing is an analysis of grade changes from first to second semester. There was no difference between the groups in raw grade change from first to second semester [$t(124) = 0.77$, $p < .44$]. If one takes a more detailed look at these grade changes (Table 1), one can see that in the X section there was a total of 40 changes, of which we called 17 large changes (more than a third of a grade). In the control section there were 39 changes, of which 18 were large. Even the distribution of good and bad changes was the same. The similarity between the two sections is simply astonishing.

At first sight, the preponderance of changes to a lower grade is most disturbing. However, remember that the X section and its control are artificially high in A/A$^-$ students, and there is no way for an A grade to change except to drop. We faced a classic “right-wall” or “ceiling-effect” problem (*8*) and could not—or at least should not—have expected any other result. Imagine a limiting case in which all our students had perfect scores in the first semester and therefore all were exactly at the limit of the right wall of the distribution. In such a class *all* grade changes must be to the downside. There can be no improvements. The preponderance of downward grade changes in the X section and the control group is the result of our starting grade distribution being too close to the right wall. The overall grade distribution in the first and second semesters was more or less the same and not significantly different from that in previous years (see Fig. 3).

One difficulty we discovered was that the small groups are remarkably powerful problem-solving machines. More than once we were fooled by the *group*’s ability to solve a problem into believing that each *member* of the group “got it”. In fact, the groups are far better than even the best member of the group. There were also a few social issues in the groups. We expected problems with reticence, and had both warned the students against it and been alert for it. We had also attempted to help students get used to the intellectual give and take by running group problem-solving classes in the first-semester recitation sessions. Still, one must admit that it is not an easy thing to put one’s thoughts out on the table, especially if there is a fair likelihood that one’s idea will be wrong. No amount of explaining that we all do this in the real world of chemistry can make it easy. There was another lurking problem of the opposite kind. Not every “A” student is suited to this approach. We were unable to avoid a certain amount of racing to the answer by some people with little regard for slower students. We were able to eliminate some of this activity by talking to the students, but not all.

What Happened in the Second Year

We learned much from our first efforts and made a number of small but helpful changes. First of all, everyone, both students and teachers, agreed that four times a week for the quotidian 50 minutes was a poor schedule. The relentless pace makes it hard for students to do the necessary reading

Table 1. Grade Changes from 1st to 2nd Semester 1998

Group	All Changes			Large Changes		
	Total	Up	Down	Total	Up	Down
X	40	11	29	17	3	14
Control	39	12	27	18	3	15

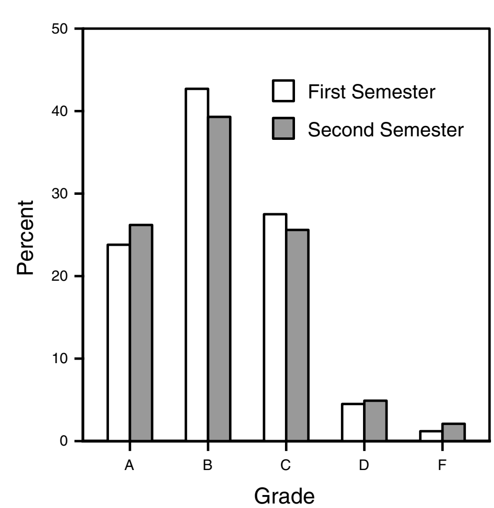

Figure 3. Overall grade distributions for the two-semester course.

in a serious way. They were pressed for time and found it increasingly difficult to keep up as the semester wore on. In the second year we switched to Monday–Wednesday–Friday for 65 minutes. The day off helped a lot, and 65 minutes was far superior to 50. The standard class period flies by, and the extra 15 minutes many times allowed for a much more satisfactory resolution of the day's efforts than was possible the first year.

We also had had a year's experience in resolving social issues and were on the lookout for dysfunctional groups so as to catch problems early. In the second year we were especially alert to the necessity of having well-working groups and advertised up front that neither reticence nor braggadocio was welcome. We were much more explicit in talking to A students about what we were offering them (*not* a grade increase!). We were also better at picking problems that evolved throughout the week and faster on our feet when we were not getting our points across.

Some problems were beyond our ability to control. For example, a rearrangement of course offerings in the chemistry department (a bifurcation of the organic course) made it impossible to maintain our control group in our second year. There were just not enough students. Our experimental course also attracted an even larger proportion of the students who did well in the first semester, thus increasing our right-wall problem and causing the usual distribution problems that arise when two different organic courses cohabit in one department.

As we could not create a proper control group, we did the best we could and constructed an improper one. We generated a cryptic control group made up of randomly selected students from earlier years who had the same grades in the first-semester course as did the students in the second year of the X section. Of course these virtual students had taken different exams at different times from the 1999 X group, so the data are surely somewhat suspect. Nonetheless, we see again that there was little difference between the experimental section and the "control" (Table 2).

Perhaps, just perhaps, we did a bit better in our second year in terms of grade changes: 1998, 28% up (control 31% up); 1999, 33% up (control 25% up). Remember, in 1999 the better class actually made it harder for us to raise grades, as more students started with high grades.

Student Evaluation

We distributed a special evaluation form for this course in addition to the usual university evaluation forms. Table 3 shows a summary of the quantitative data gathered from the 50 responses from the total of 63 students in 1998 and 52 responses from the 58 students in 1999. From the students' point of view, even though the grades didn't change in a positive fashion overall, the experiment was a great success, especially in the second year. Note that nearly every evaluation category improved the second time through. In the narrative section of the university evaluation and in the "other comments" section of our evaluation, many students in both years stressed the "intangibles". Overall, they thought that the cooperative format was fun and that they got more out of the class than just facts or even just a grade. (And we may have had some successes in raising grades, even with our severe right-wall problem.)

The question of what we have to offer to a good student is an interesting one that gets to the heart of what we are trying to do. The numbers are telling us that we are not going to change grades in a serious way, at least until we can teach the whole class and avoid our right-wall problem. What we do have to offer, we think, is an opportunity to understand the material in a deeper way. What we offer the A student is the chance to learn better and more deeply by helping to teach the material to others. We all know how quickly a shallow understanding is uncovered when we try to explain to someone else something we don't really know in depth. We may survive the class, but we are often sent scurrying off to the literature by such an experience. We told the A students this, and tried to encourage them to alert us when the inevitable happened to them in the course of an explanation. Some got the idea, a few did not.

Table 2. Grade Changes from 1st to 2nd Semester 1999

Group	All Changes			Large Changes		
	Total	Up	Down	Total	Up	Down
X	33	11	22	14	3	11
Control	28	7	21	10	2	8

Table 3. Student Evaluations of the X Course

Question	Response	% of Responses	
		1998	1999
I did the reading before class	Always	26	17
	Often	38	63
	Sometimes	26	15
	Rarely	10	4
I came to class	Always	60	56
	Often	36	42
	Sometimes	4	2
	Rarely	0	0
Percentage of classes missed	0	9	11
	1–3	16	11
	5	13	15
	10	9	11
	15–20	2	1
	40–50	1	1
Has this style of learning had a positive effect on your attitude about science?	Yes	67	83
	No	13	10
	Unsure	20	7
Has this style of learning had a positive effect on your attitude about chemistry?	Yes	70	83
	No	15	10
	Unsure	15	7
Overall, what is your attitude towards cooperative learning?	Positive	74	75
	Negative	24	6
	Indifferent	2	19
Did this class increase your ability to think independently?	Yes	63	83
	No	35	15
	Indifferent	2	2
Did this class increase your ability to communicate effectively with others in a scientific manner?	Yes	78	77
	No	22	19
	Indifferent	0	4
How effective was the cooperative learning style in developing your understanding of organic chemistry?	Very effective	27	46
	Somewhat effective	53	33
	Somewhat ineffective	6	4
	Very ineffective	12	15
	No effect	2	2
Overall, was this class a rewarding experience?	Yes	76	88
	No	18	10
	Don't know	6	2

Our Evaluation

This class was much more work than the regular section. Moreover, it was far more stressful for us, especially in the first year. The wear-and-tear quotient was very high. Some of the stress arose from "good" reasons. We got to know the students much better than we did in the large lecture course, and it really hurt when they didn't do well, especially since they volunteered for this experiment. The worst aspect is dealing with students who did not do as well as they had in the regular first semester. It is easy to feel responsible for hurting a student with this experiment. One gets personally involved. Statistics say we did not hurt them, but statistics are not individuals, and there were three or four students each year who did not prosper in the experimental section. Although there were others who flourished, somehow one feels the pain of the failures more than the joy of the successes. But there are other, "bad", reasons for the stress we felt. Students these days often seem to resent not getting a high grade. One can imagine many reasons. For example, grade inflation has not seriously reached the organic course here, and students are unhappy with the C's that we still give, especially in contrast to the perception that relatively easy A's are available in summer courses or in other parts of the university.

In the first year of this experiment, one of us (MJJr) was also teaching the regular lecture course, and it was easy to see that the presence of the experimental section made some students in the regular course angry. Some suspected or felt certain that special advantage was given to the X class "because Jones wants the experiment to succeed." In a few cases, suspicion led to real paranoia or blindness. For example, we got complaints on the regular course evaluations that there was no access to the "special" problems given out in the experimental section. In fact, all problems and their answers were posted for everyone on the Web, and that fact was clearly announced more than once in the lecture course. There was also suspicion that we were slanting exam questions to favor the experimental section. In fact, we leaned over backward not to do this and probably were guilty of disadvantaging the X class for this reason.

Problems Remaining and What Happened in the Third Year

Having demonstrated, at least to our own satisfaction, that the problem-solving approach is for many students at least the equal of the traditional lecture-based course, we faced a number of problems. Will such a course survive in competition with the traditional lecture? Can the problem-solving approach be built into the whole organic course through expansion of the method into the first semester? Critically, can we find ways to accommodate more students in this system without increasing the costs inherent in the low student/teacher ratio?

It is apparent that many of these questions have to do with finding ways in which to make the problem-solving approach practical. In view of the low student/teacher ratio (1:20 as opposed to 1:200+ in the large lecture course) and the wear and tear on the teaching staff, a critical question for the future is whether the results justify the extra effort and costs. There are several reasons why they might. The experimental problem-centered method and the traditional lecture format are at least equally effective teaching methods for most students. As the student evaluations show and as we can attest, despite the hard work involved for all, it is the former that is likely to be more interesting—and more fun—for both students and faculty. Such advantages seem sure to translate into better overall long-term learning.

In the third year of this experiment we expanded to the first semester and allowed increased enrollment to 96.[3] We opened with lectures, using them to build the structural and conceptual grammar of organic chemistry, and gradually worked in problem-solving sessions. By the end of the semester we were entirely in the problem-solving mode, which ultimately accounted for 50% of the class time. The second semester again ran essentially entirely in the "X style", the few lectures being restricted to mornings after examinations and to special topics.

Now we faced the severe problems of logistics and costs. With approximately 25 groups we were spread very thin in the fall. We decided to try to cut costs and increase coverage by enlisting undergraduate alumni of the course as teaching assistants in the second semester.[4] University regulations allow this procedure as long as the undergraduates are not involved in grading. Such a scheme is certainly cost-effective, as undergraduates are paid only a modest sum and there are no graduate student tuitions or stipends to be paid. We selected student alumni of the first two years of the experimental course on two bases, command of the material and a perceived ability to relate well to others. We had seen these students work in their groups and were confident we could select a corps of successful young teachers. The undergraduates "cruised" the room as did we, and the whole teaching group met once a week for lunch to discuss problems and opportunities.

The results surpassed our expectations. Our highly subjective "feel" of the room told us that things were going well during the semester. Because of the regrettable lack of a control, we did not subject the results to quantitative analysis, but there were no obvious differences from the first two semesters. The student evaluations told a remarkable story. The students were asked to rate the performance of the undergraduate teaching assistants. The results were overwhelmingly positive, as shown by the tabulated responses (Table 4).

Of course, there are other benefits to this scheme. The undergraduate teachers themselves are not only learning the material in a new way as they teach it, but they are exposed to the other side of teaching—both the joys and the frustrations. The result was more than one changed notion of the future in our first group of undergraduate teaching assistants.

Table 4. Student Ratings of Undergraduate Teaching Assistants

Highly Effective		Effective		Somewhat Ineffective		Very Ineffective	
No.	%	No.	%	No.	%	No.	%
17	37	22	48	3	6.5	4	8.5

The Future

We have surely learned that there is a very strong market for this approach. Despite its experimental nature, despite minimal advertising and a very early morning hour, despite our warnings that this class expects a great deal of effort from students, we attract more than half of the available students, and the demand for this course has increased every year. This year we start with 133 students, up more than 33% from last year. We begin the fourth year of this experiment with no graduate student assistants, but with a crew of veteran and new undergraduate teaching assistants. Although we hope to report further on these difficult matters, one can follow the progress of the 301X and 302X courses on their Web pages, available through *http://www.princeton.edu/~mjjr/*, where all the materials we use can also be found.

Acknowledgments

AZB thanks Princeton's Council on Science and Technology and The Camille and Henry Dreyfus Foundation Special Grants in the Chemical Sciences Program for support. We all appreciate helpful conversations with Grant Krow of Temple University. Mary W. Baum and George McLendon also made a number of helpful suggestions.

Notes

1. We describe the situation as of the start of our experiment. There are now two organic courses: the ongoing experimental section and a more bio-oriented traditional lecture course.

2. The three were on the evolution of the Cope rearrangement from 1,5-hexadiene to bullvalene, the 2-norbornyl cation, and the Fischer proof of the structure of glucose.

3. In the fourth year, 2000–2001, we allowed open enrollment in the two courses. Approximately 60% of the students opted for the experimental course.

4. This solution was suggested to us by George McLendon. Our thanks—and those of the undergraduates involved—go to him.

Literature Cited

1. Schearer, W. R. *J. Chem. Educ.* **1988,** *65*, 133. Birk, J. P.; Foster, J. *J. Chem. Educ.* **1993,** *70*, 180. Ellis, A. B. *Chemtech* **1995,** 15 and references therein. Felder, R. M. *J. Chem. Educ.* **1996,** *73*, 832. Dougherty, R. C. *J. Chem. Educ.* **1997**, *74*, 722. Robinson, W. R. *J. Chem. Educ.* **1997**, *74*, 622, 623. Kogut, L. S. *J. Chem. Educ.* **1997**, *74*, 720. Moog, R. S.; Farrell, J. J. *Chemistry: A Guided Inquiry*; Wiley: New York, 1997. Towns, M. H. *J. Chem. Educ.* **1998,** *75*, 67. Spencer, J. N. *J. Chem. Educ*. **1999,** *76*, 566.
2. Farrell, J. J.; Moog, R. S.; Spencer, J. N. *J. Chem. Educ.* **1999,** *76*, 570.
3. Gosser, D. K. Jr.; Roth, V. *J. Chem Educ*. **1998,** *75*, 185. Gosser, D.; Roth, V.; Gafney, L.; Kampmeier, J.; Strozak, V.; Varma-Nelson, P.; Radel, S.; Weiner, M. *Chem. Educator* **1996,** *1*, 1 (*http://www3.springer-ny.com/chedr*). Kampmeier, J. A.; Varma-Nelson, P.; Wegegaertner, D. K. *Peer-Led Team Learning. Organic Chemistry*; Prentice-Hall: New York, 1999.
4. We know of only a few reports on cooperative learning in the organic lecture course. Dinan, F. J.; Frydrychowski, V. A. *J. Chem. Educ.* **1995,** *72*, 429. Paulson, D. R. *J. Chem. Educ.* **1999,** *76*, 1136. We have been helped greatly by conversations with Grant Krow at Temple about his impressive, though unpublished, efforts in this area.
5. Loudon, M. L. *Organic Chemistry*; Benjamin/Cummings: Redwood City, CA, 1995.
6. Jones, M. Jr. *Organic Chemistry*, 2nd ed.; Norton: New York, 2000.
7. See Gardner, H.; Hirsch, E. D. Two Views on How to Get Johnny to Read and Think; *The New York Times*, Sep 4, 1999, p A12. Gardner, H. *The Disciplined Mind: What All Students Should Understand*; Simon and Schuster: New York, 1999.
8. Kazdin, A. E. *Research Design in Clinical Psychology*, 3rd ed.; Allyn and Bacon: Needham Heights, MA, 1998. For a marvelous examination of more profound right- and left-wall problems, see Gould, S. J. *Full House*; Harmony: New York, 1996.