Assessing the Conversational Classroom

William M. Waite
Michele H. Jackson
Amer Diwan
University of Colorado
Boulder, CO 80309
{William.Waite,Michele.Jackson,Amer.Diwan}@Colorado.edu

ABSTRACT

Conventional wisdom states that the large lecture class is an efficient way to teach a common core of material to large numbers of introductory students. We questioned that model, and chose a required second-year course with enrollments in the 60-120 range to test a classroom strategy that dispenses with the lecture entirely. Instead, each class session consists of a conversation among the students and professor. A nine-semester longitudinal study shows that this change results in a sustained improvement in student performance.

Categories and Subject Descriptors

K.3.2 [**Computers and Education**]: Computer science education

General Terms

Design, experimentation, performance, human factors

Keywords

Curriculum issues, course pedagogy, classroom management, CS educational research, communication skills

1. INTRODUCTION

The people who hire our graduates consistently characterize them as "technically competent, but unable to work in groups". This remained true even after we had increased the number of project courses and the opportunities for teamwork.

Because these structural solutions had failed, we undertook an extensive study of student culture [16] in an attempt to get a better understanding of our target audience. We used a process of ethnographic observation and in-depth, semi-structured interviews of students in three undergraduate courses over two and a half years. One of the first things that we noted from this data was a widespread passivity. Here's how one student characterized their model of education:

> ... all through elementary school and college you're taught that you're supposed to sit around this oracle of knowledge and the oracle will impart the knowledge upon you. ... Unfortunately, people are terrified of contradicting the oracle, and think people who do so are stupid. They don't want to be stupid.

Studies have shown that classrooms in science and engineering often perpetuate this attitude [14]. Changing this situation requires a significant cultural shift for both students and faculty [7], because university science and engineering departments believe that large lecture classes are an efficient way to teach a common core of material to large numbers of introductory students.

In contradistinction to this lecture-based "oracle" approach, we believe that education should be a collaborative process involving interaction among the students as well as with the faculty. If they have this cultural orientation, graduates should be able to perform competently in groups when they enter the work force. We therefore decided to try changing the classroom strategy for a large, required, second-year course taken by students in both Electrical Engineering and Computer Science. The new strategy both increased the amount of "active learning" (in-class discussion and problem solving) and transferred control of the class session to the students [15]. The result was a significant improvement in academic performance that has been sustained for four years.

There are a number of challenges associated with attempting to assess the effectiveness of an emergent and unpredictable intervention. We believe that the data collection and analysis methods we used for this experiment are important in an era of increasing demand for accountability from educational institutions. These techniques enable us to answer certain kinds of questions from students, administrators, and our colleagues about course quality. They also provide information for accreditation reviews [8].

Section 2 describes the characteristics of the course and Section 3 contrasts the two classroom strategies we used. Our data collection and analysis procedure is detailed in Section 4, and we discuss the results in Section 5. Finally, Section 6 draws some conclusions.

2. THE COURSE

ECEN 2120 is a second-year course required of all Electrical Engineering majors and Computer Science majors. Its purpose is to introduce the students to machine architecture and assembly language, and to provide them with a basic understanding of how a computer can be used as a component of a larger system.

This course carries five credit hours. Students meet in the classroom for three 50-minute periods and in the laboratory for two 110-minute periods each week. Laboratory sections are limited to 20 students. We can schedule up to six laboratory sessions per day, with either two or four days per week, depending on the enrollment. Thus the maximum number of students that can take the course in any given semester is 240.

Students are assessed primarily by means of three examinations. We divide the semester into thirds with two hour-exams, each of which contributes 15% of the total points. The final exam is worth 40%, and the remaining 30% is divided among the weekly laboratory exercises. These vary in weight with the complexity of the exercise, increasing as the semester progresses. The last exercise accounts for about 3% of the total points.

In the first third of the semester, students explore the characteristics of the central processor and learn its assembly language. Interaction with a variety of peripheral devices, interrupt handling, and mechanisms for dealing with concurrency occupy the middle third. The remainder of the semester covers methods for introducing wall-clock time into programs and for managing memory, and some special topic of interest to the students and the professor.

The sequence of laboratory exercises effectively determines the content of the course. Those exercises have been essentially unchanged since the course was first offered in the fall of 1995. Even when we made the transition from the Intel to the Motorola architecture in the fall of 2003, only a few slight alterations were required.

Nominally, students work on laboratory exercises in pairs. The teaching assistant has the option of allowing singles or groups of three in exceptional circumstances. We have found that a group of four or more is invariably dysfunctional in the context of this course.

All of the laboratory exercises are available on the Web at the beginning of the semester [1]. The first exercise familiarizes the students with the lab computers and basic procedures. Each exercise thereafter requires that a program be developed, and encompasses two sessions. A written assignment is handed in at the beginning of the first session, and the second is devoted to student demonstrations of their programs.

The written assignment due at the beginning of the first session attempts to focus the students' attention on the important concepts of the current program, and to expose possible misunderstandings. The first part of this session is devoted to student questions, with the remainder used to work on the program. If the students don't complete the program during that laboratory session, they may return whenever the laboratory is not specifically scheduled.

Each group has five minutes to demonstrate their program to the teaching assistant during the second laboratory session, and the teaching assistant then has a further five minutes to question them about what they have done. The teaching assistant selects the demonstrator at random, which means that all authors must understand their program. Similarly, questions may be directed at random to any author. This approach to assessment means that simply copying a solution from last semester, or from a fellow student, is not sufficient. We believe that if a student actually understands the program, and can convince the teaching assistant of that fact, then they have gained the knowledge they need.

The course material is broken down along a detailed *course time line* that associates topics with specific classroom sessions and textbook sections [1]. This time line represents a reasonable schedule for students to assimilate the information, assuming that they adhere to the recommended practice of devoting three hours per week per credit to a class; the 15 hours per week for this particular course includes the scheduled contact hours.

3. USE OF CLASS SESSIONS

One of us (Waite) was the lead developer for ECEN 2120, and has taught the course nine times. He used a lecture format for the class sessions from 1995 through 2001: after first calling for and answering any questions from the students, he would explain aspects of the material listed on the course time line for that class session or develop sample code illustrating techniques to be used in the laboratory.

We modified our classroom strategy for ECEN 2120 to create an environment resembling an engaged, intellectual conversation [15]. This conversation was not informal or casual, but was closer to the interaction typical of an upper-division seminar in which participants learn through a process of engaging other people's ideas, surfacing issues, experiencing surprise, and reflecting on their own ideas [13]. This is a form of learning that cannot exist without conversation.

The conversational strategy also exposes the *process* of thinking about course material. Participants see how others reason, in addition to being exposed to a variety of possible conclusions. They learn how to make useful judgments more quickly than do people studying the same material individually, and the judgments of the group are usually better than those of the individual [3].

A conversational class would begin, as before, with Prof. Waite asking whether there were questions. Instead of simply answering those questions, however, he would solicit answers from the other students. If there were no volunteers, he might pose a simpler question and try to get the class to use that as a stepping stone to the original answer. Ultimately the original question would get answered, but the answer usually brought a host of related insights.

When programming problems came up in the course of the discussion, the students would often take 10 or 15 minutes to work on them with their neighbors. During that time, Prof. Waite would circulate through the room answering questions and discussing issues with individual students. When everyone seemed to have reached a stopping point someone would go to the board and explain their solution, usually resulting in a lively discussion of various ways to solve the problem and the merits or otherwise of each. The lack of a single "right" answer was distressing to some students at first, but they eventually came to understand that a problem may have many solutions and that design is usually a matter of compromise.

4. DATA COLLECTION AND ANALYSIS

Any collaboration requires that participants yield authority over their work to others, and that those others accept the yielded authority [6]. In the conversational classroom, the professor yields authority over the use of class time to the students. Of course this requires the students to accept that authority by studying the appropriate material before class and coming prepared to direct the group's attention to the important points.

Assessment of a new way of interacting with the students requires methods (such as interviews) that allow us to be sure we capture student attitudes and behaviors and also methods (such as analyses of student performance) that can provide a more objective or impartial assessment of our techniques. Careful collection of these different types of data was critical to the formative and summative assessment of this project. Finally, we attempted to build in assessment strategies that could allow us to be surprised or to find things we did not expect or anticipate. This would force us to develop our model more fully, or to reassess our explanations and expectations, or to design new intervention strategies.

4.1 Overall Grades

A student receives a numeric grade on each written assignment and each demonstration (Section 2), on each hour exam, and on the final. These numeric grades are entered into a program called *Gradekeeper* [2] that weights them appropriately and sums the weighted values to yield a single number for the course to date. Gradekeeper produces reports for both professor and students. The student reports are posted weekly on the course web site, using a random identifier generated for each student at the beginning of the semester.

The numeric grade distribution is very noisy, preserving small differences that are due to random factors. The grades tend to cluster, however, indicating that a number of students have approximately the same level of competence. Letter grades for the course are not assigned on the basis of cutoff values determined a priori, but rather on the basis of this clustering: one Gradekeeper report for the professor lists the students in rank order of weighted points. At the end of the semester, Prof. Waite determines tentative divisions between letter grades by looking for the largest first differences in this rank order. These tentative divisions may be revised after discussion with the teaching assistants about the students at the boundary, but usually there are no changes.

Table 1 shows the number of students in the class and the fraction receiving each letter grade for the nine semesters covered by this study. Here we have further aggregated the results, combining grades with + and – modifiers into the group for the letter they modify (i.e. all students with grades of B–, B, and B+ are considered as "B students"). In addition, the tables show the relationship between numeric and letter grades, giving the highest and lowest numeric scores corresponding to each letter grade.

Each bar in Figure 1 was obtained by dividing the total number of students receiving a particular grade under one of the two classroom strategies by the total number of students taught under that strategy. It clearly shows a shift from C through B to A associated with the four semesters in which the conversational classroom was used; what is less certain is the source of this improvement.

When we evaluated the results, we were frankly astounded. We simply could not credit that a simple change in handling classroom interaction could produce a percentage of students getting A or A– so far above the historic norm, and spent a lot of time and effort exploring other possible causes [15]. In the end, however, we were forced to admit that everything but the classroom strategy was unchanged from the previous five semesters.

There have been no demographic or curricular changes since 2002 that could have affected our students, and for the most part the exams have remained at the same level of difficulty. (The one exception is considered in Section 4.2.) The assignment of letter grades, however, is carried out independently each semester. We therefore examined the cutoff points listed in Table 1 for any systematic differences.

We averaged each pair (lowest x,highest y) from the tables to obtain the cutoff point between grades x and y. Our null hypothesis was that these cutoff points were drawn from the same population for each classroom strategy. F-tests showed that, for each pair of adjacent grades, the variances in the cutoff points for the two strategies could have come from the same distribution. t-tests did not indicate any statistically-significant difference between the means in the standard lecture and the conversational classroom data for any of the corresponding cutoff points. Thus we concluded that the grade improvement shown in Figure 1 is not due to a change in cutoff points.

Finally, we investigated the statistical significance of the difference in the percentage of students getting particular grades between the lecture strategy and the conversational classroom. As Figure 1 implies, there was no significantly different percentage of B, D, and F students. The variances of the C-student samples for the two techniques could have come from the same distribution, and a t-test shows that the difference between the two means is significant at the 3% level.

The F-test tells us that the variances of the A-student samples are significantly different between the techniques. Unfortunately, the number of samples is too small to use the t-test reliably under these conditions, so we cannot conclude that the difference in the proportion of A students is statistically significant. Nevertheless, because of the significant reduction in the proportion of C students and the fact that proportions of students with other grades remained the same, we feel justified in concluding that the conversational classroom *does* improve student performance over a standard lecture presentation.

4.2 Exam Grades

Table 1 shows that the mean examination grades are relatively consistent over a single class, with the second exam usually showing the best performance. The per-class variances of these means have similar values, with the exception of those for 2002 and 2004. These are more than double the next largest value, and that led us to look more carefully at the details of Exam 2 in 2002 and Exam 1 in 2004.

When we analyzed the initial data for the conversational classroom in the summer of 2002 [15], we felt that the anomalous results on Exam 2 might really be the underlying cause of the increase in the percentage of As. This turned out not to be the case. We determined the effect of that exam on the final grades for the semester by "excusing" all of the students from it and then re-allocating the letter grades on the basis of the resulting rank order.

When a student is excused from a graded item, the Gradekeeper program distributes the weight of that item over the other items of the same class (in this case the other examinations). The weight is distributed in proportion to the weights of the other items. Roughly speaking, the effect of excusing a student from an item is the same as assuming that they completed that item with *their* average performance on items of that type.

The final grades for 2002 were unaffected when we excused the students from Exam 2. Only two students would have had their final letter grades change: one went from a B+ to

Table 1: Collected Results

	Lecture					Conversation			
Year	1995	1997	1999	2000	2001	2002	2003	2004	2005
Number of Students	93	118	136	125	83	116	82	73	66
Students with A	17%	14%	15%	14%	16%	35%	24%	15%	21%
Students with B	42%	23%	36%	18%	12%	24%	28%	41%	27%
Students with C	30%	54%	38%	49%	42%	31%	27%	23%	33%
Students with D	9%	5%	7%	14%	20%	4%	11%	12%	14%
Students with F	2%	4%	4%	6%	11%	6%	10%	8%	5%
Highest A	93%	91%	95%	88%	95%	95%	94%	94%	91%
Lowest A	77%	73%	79%	76%	76%	75%	79%	82%	72%
Highest B	75%	71%	77%	75%	73%	74%	77%	81%	71%
Lowest B	61%	63%	66%	67%	68%	65%	63%	66%	61%
Highest C	59%	62%	65%	66%	66%	64%	61%	65%	60%
Lowest C	45%	41%	51%	48%	53%	45%	47%	49%	47%
Highest D	44%	40%	50%	46%	52%	42%	45%	46%	45%
Lowest D	36%	35%	47%	37%	38%	36%	37%	40%	36%
Highest F	31%	35%	45%	35%	34%	33%	34%	36%	31%
Lowest F	27%	24%	37%	20%	28%	20%	19%	20%	19%
Exam 1 Mean	57%	65%	57%	45%	52%	55%	50%	40%	56%
Exam 2 Mean	65%	52%	59%	54%	50%	75%	60%	63%	43%
Final Exam Mean	54%	50%	52%	49%	52%	56%	56%	59%	52%

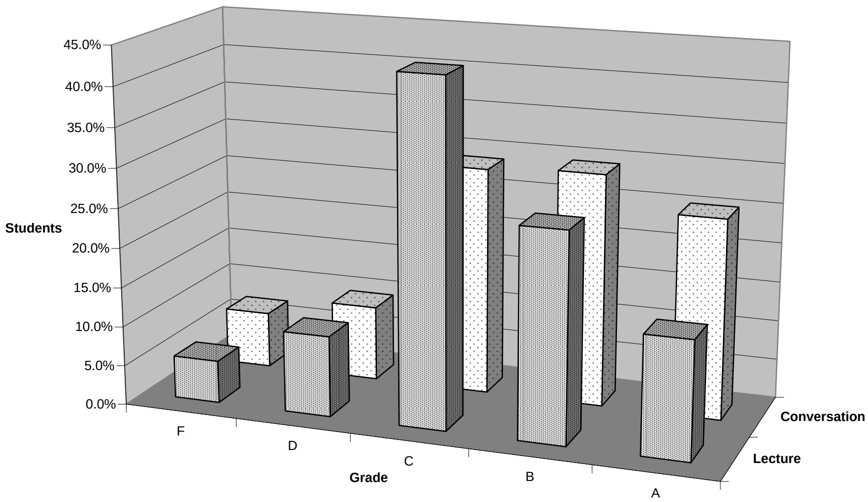

Figure 1: The Shift in Grade Distribution.

Table 2: Ignoring a Broken Exam

Year	2002	2003	2004	2005
Number of Students	116	82	73	66
Students with A	35%	24%	23%	21%
Students with B	24%	28%	22%	27%
Students with C	32%	26%	20%	33%
Students with D	3%	11%	4%	14%
Students with F	6%	11%	4%	5%
Highest A	95%	94%	97%	91%
Lowest A	75%	79%	79%	72%
Highest B	74%	77%	77%	70%
Lowest B	65%	63%	66%	61%
Highest C	64%	61%	65%	60%
Lowest C	45%	47%	43%	47%
Highest D	42%	45%	41%	45%
Lowest D	36%	35%	36%	36%
Highest F	33%	34%	28%	31%
Lowest F	20%	19%	19%	19%
Exam 1 Mean	55%	50%	—	56%
Exam 2 Mean	75%	60%	63%	43%
Final Exam Mean	54%	56%	59%	52%

an A– and the other from an A– to a B+.

We also asked several people both inside and outside the university to rate all of the exams in terms of difficulty. The exam in question was ranked near the middle of its group; see [15] for more details of this process. Thus we have no explanation for the excellent performance on Exam 2 in 2002.

There is considerable evidence, however, that Exam 1 from 2004 was poorly conceived:

- Normally the teaching assistants are asked to assess the exam before it is given. In this case, an unfortunate series of miscues prevented them from doing more than a cursory scan.

- The target machine for the course changed from an Intel to a Motorola architecture between 2003 and 2004. Because the Motorola architecture is more regular, many of the types of questions normally asked on Exam 1 were irrelevant. Previous exams were thus an unreliable study guide (previous exams are available on the web [1], and students are explicitly encouraged to look at them).

- A larger fraction of the exam questions than usual involved writing code. We have seen over the years that mean scores are lowest for this type of question.

We therefore went through the same process of "excusing" the 2004 students from Exam 1 and re-determining final grades. The story this time was quite different from that of 2002: 42 of the 73 grades changed, with only two going down. 12 of those changes were from B's to A's. Table 2 shows the complete results for the conversational classroom approach with these changes.

The variances of the A-student samples in Table 2 are still too different for a reliable test of significance, but it seems clear that the basic conclusions remain unchanged.

4.3 Student Comments

Near the end of the semester, students fill out questionnaires in which they can comment about what they felt were the most successful and least successful aspects of the course. Table 3 summarizes the responses involving lecture style. For example, 68 of the 93 students returned questionnaires in 1995, a 73% response rate. Of these 68, 11 listed the lectures in the "most effective aspects" block or made a comment praising them; 5 listed the lectures in the "least effective aspects" block or made a comment denouncing them. The last two rows give the fraction of the responses that comment on the lectures, and of those the fraction in which the comment was positive.

We did not discuss our approach with the students when we began using the conversational classroom strategy in 2002. Table 3 shows that the number of comments about the lectures was approximately the same as usual, and perhaps a bit low. The number of positive comments was also in line with earlier semesters. From 2003 on, however, we made a point of explaining the strategy in the first lecture. The fraction of returned questionnaires making comments about the lectures doubled, and half or fewer were favorable.

The questionnaires are administered by one of the students during the last 20 minutes of a regular class. Although the last week of class is the mandated time for administration, the actual date is left to the professor. Prof. Waite never announces the date ahead of time; on a convenient day he simply requests that the students fill out the questionnaire, states that he believes that the comments are the most valuable part, and returns to his office. The student administrator then hands out the questionnaires, provides some additional instructions and assurances of privacy, collects the completed material, and delivers it to the Dean's office. Very few students leave without completing the questionnaire, and thus the response rate serves as a rough indicator of classroom attendance.

It appears that attendance was low in 2003 and 2005, but in accordance with historical norms in 2002 and 2004. As noted above, the students seem to have regarded 2002 as a "normal" year, even though the conversational classroom strategy was being used.

In 2004 we used "clickers" [9], devices that allow students to answer multiple-choice questions during the class period. (Clickers are most commonly used to implement Mazur's "concept tests" [11].) One well-documented effect of using clickers in a class is increased attendance [10], so we assume that attendance in 2004 would have been similar to that in 2003 and 2005 if clickers hadn't been used.

5. DISCUSSION

The research literature in computer science education provides examples of a number of mechanisms for assessing pedagogical changes. Here we use letter grades, the evaluations that go forward with our students, as acceptable measures.

We believe that the conversational classroom goes further in the direction of explicitly encouraging the students to take responsibility for their education than most current initiatives. Such a shift in responsibility tends to occur in any case as the student matures, but our experiment shows that it can be accelerated by a change in classroom culture.

Finally, our results address the question of class attendance. How important is it? Is it worth trying to force

Table 3: Student Comments

Year	1995	1997	1999	2000	2001	2002	2003	2004	2005
Number of Students	93	118	136	125	83	116	82	73	66
Responses	68	103	92	88	58	78	46	64	35
Lectures best aspect	11	20	24	16	17	15	10	16	9
Lectures worst aspect	5	7	7	13	0	2	13	24	9
Response rate	73%	87%	68%	70%	70%	67%	56%	88%	53%
Lecture comments	24%	26%	34%	33%	29%	22%	50%	63%	51%
Lectures best aspect	69%	74%	77%	55%	100%	88%	44%	40%	50%

attendance when many students are trying to reduce their fixed time commitments, and may be willing to trade off a grade to do that? We know that different students have very different reasons for attending particular courses [12], not all of which involve the level of involvement desired or envisaged by the professor.

5.1 Grades as Course Assessment Tools

Some people view the assessment of a course on the basis of the grades given as subjective and not convincing. This is an interesting position, and a significant indictment of the whole process of student evaluation. The problem, of course, is a lack of standardization in grading across particular curricula, departments, universities, and higher education generally. We understand these problems, but we see no other objective measures that can be used in an arbitrary situation; standardized tests can be developed and applied across larger populations than a single class, but the effort involved ensures that these will be limited to a few very common courses.

Using grades as an assessment tool requires us to be concerned not only with consistency between content and assessment, but also with consistency across terms. We need to be aware that our grading is not always neutral, and that there may be contingencies that change the nature of the instrument. At the same time, this cannot become an excuse for not getting the expected result. We addressed these issues by extensive use of student feedback, observation, and external analysis of testing instruments.

ECEN 2120 is intended to provide students with an operational understanding of how a computer can be used as a component in a larger system. We assess that understanding in the laboratory by having the students actually build such systems, and on examinations by asking a mixture of questions representing all levels of Bloom's taxonomy [5]. All of the laboratory exercises are on line, as are all of the examinations given in the course to date. Almost all of the questions on a given examination are either identical to questions on previous examinations, or assess the same knowledge in a similar fashion. (As discussed in Section 4.2, the first exam in 2002 had a large number of new questions because of the change in machine architectures.) Thus we believe that our basic assessment tools come as close as possible to standardized instruments for this course.

Our process for arriving at an overall grade for the course (see Section 4.1) filters out most of the noise associated with the primary results. The clustering process that we use to determine the cutoff points accounts for small variations in instruction and grading from semester to semester. Table 1 shows that this process does not bias the overall letter grades.

The university's grading policy is defined as follows:

> The following grading system is standardized for all colleges and schools of the university. Each instructor is responsible for determining the requirements for a course and for assigning grades on the basis of those requirements.
>
> **A** superior/excellent
>
> **B** good/better than average
>
> **C** competent/average
>
> **D** minimum passing
>
> **F** failing

We assert that these grades can provide the basis for course assessment in experiments involving pedagogical techniques. Indeed, if assessment of course pedagogy cannot rest on grades then how can we claim that we have assigned those grades "on the basis of [the course] requirements"?

5.2 Active Learning and Responsibility

Over the past few years there has been a significant shift in a number of disciplines (including mathematics and physics): less lecturing and more in-class discussion and problem solving. The discussion and problem solving usually takes place among students, however, with the professor deciding on the order of the activities and the amount of time to be spent on each. Questions are posed by the professor, and ultimately the professor provides the "right" answers.

One of the cornerstones of our approach, however, is to place the choice of questions and the amount of time spent on each in the hands of the students.

In our model, the professor is responsible for what might be termed an "agenda" for the course: a detailed syllabus defining the material for which the students will be responsible and providing a time line by associating particular topics with particular days. This syllabus has the characteristics of any good agenda:

- It is clear
- It is consistent with the goals of the course
- It can be accomplished reasonably, given the resources available

Thus each class meeting has particular objectives — we know what we are trying to accomplish in that meeting.

The student is responsible for becoming familiar with the material to be discussed at each meeting, and participating in the conversation. This student's comment on the 2004 course questionnaire indicated that they understood the model:

> I really like how the first question he asks is "Are there any questions?" — we aren't being lectured at, we're participating in the conversation. The labs were also extremely helpful in understanding the material. However, this approach puts a lot of responsibility on the student. I think this is good because it forces you to work — in other words, the grade you get is determined by the effort you put into it.

This 2004 student, however, did not want the responsibility:

> In all decency, sir, you must present all the relevant, testable material in class! How can you expect us to teach ourselves subject matter that is as foreign to us as it is complex? *Lecture*, do not "discuss". Discussions help a small portion of the class; lectures provide relevant information to us all. I would have appreciated real lectures, especially prior to tests.

These two comments succinctly characterize the pros and cons of the conversational classroom, as well as the challenges of this particular course. The material *is* complex and unfamiliar. It isn't easy for a novice to absorb, no matter how it is presented. This is why the course carries five credits and involves both classroom and laboratory experience.

Students such as the one quoted above may well be expressing anger about the lack of a standard lecture because they have discovered that they can do well in many courses by simply reproducing what the professor says, without really gaining a deep understanding of the concepts involved. If they are unable to accept responsibility, they will not gain any benefit from a conversational classroom. They don't believe (or don't care) that they are effectively cheating themselves of their education by requesting the answers to the test instead of putting forth the effort to master the material.

A challenge here is that ethnographic interviews essentially replace the oracle of the instructor with the oracle of the student. We need to assess the effectiveness of the intervention independent of the bias of particular students. We chose to address this by gathering a large number of interviews so that we could rely on "patterns" of student responses, and by using a rich theoretical model of interaction and communication [13, 18] to guide our interpretation of those responses.

5.3 Attendance

From 2003 on, we used the first lecture to explain our view of education as a collaborative endeavor, and how each of the components of the course fit that model. We pointed out the students' responsibilities, and the benefits that they could expect by finding out what it was that they didn't know and didn't know that they didn't know [4]. Here's how one student put it during an interview:

> I just remember that discussion being helpful, because someone would ask a question, and I was like, wow, I actually didn't understand that as well as I thought I did.

Lower attendance under the conversational strategy could be due to students opting out of the lectures either because they didn't want to participate in the discussion or hadn't read the material. This student explained the dilemma in their interview:

> It took a lot more responsibility before the class. A couple of times I wouldn't have the time to do the reading beforehand, and in those cases the lecture was worthless to me. I might just as well have not gone. So, as long as you keep up, which in a fire hose course it's kind of difficult to do, but as long as you do, then it's very useful.

The same student went on, however, to expand on the advantage of taking the responsibility of reading and participating:

> Usually after I read the book I had some general idea of what was going on, but very, very fuzzy. After the lecture it usually solidified real well, actually to the extent that I can't remember a time where I just sat down with notes or a book and tried to really hammer on something and figure it out. I don't think I ever had to study in a traditional sense.

This view is interesting because the reference to not having to "study in a traditional sense" emphasizes that there is more than one way for a student to gain the necessary competence to pass the course. It shows the critical importance of interviews to assessment of a pedagogical technique. We were most able to be surprised, to have our expectations and theories challenged, through student voices. Students were, in effect, almost co-researchers — suggesting insights that we would not otherwise have gained and driving us to develop our model of how and why this particular classroom strategy works.

6. CONCLUSION

The conversational classroom strategy increases active learning and encourages the students to take responsibility for their own education. This represents a significant shift in classroom culture, which has several implications:

- It may be threatening to some professors who wish to maintain control over the entire process.
- Careful planning is needed before the semester begins, because the course must have a clear agenda that will allow the students to master the material within the time guidelines set by the university.
- Intensive preparation may be necessary for the class sessions as well, because the conversation may go in any of a number of directions. The professor needs to be able to contribute meaningfully, ensure that a few students don't monopolize the floor, provide examples to illustrate points as they come up, and help the students to move on when a topic is exhausted.

Further research is needed to ensure that this strategy can be transferred to other classes. A student in a third-year course where the technique was used had these answers to the questions in our standard interview protocol:

- What do you think of his use of discussion in class?

> I think it's a good idea, because he finds out what we're thinking and he tries to spread it out, other than the five or six people that always know, the A students that always seem to know the answers and have their ideas. He tries to encourage participants from the people who sit in the back of the room and don't want to say anything.

- How typical is his discussion format in Computer Science courses?

> Very rare. Most of the time it's lectures. Some of my professors read right out of the book. The closer they follow the book, the less I follow the lecture. When they're taking the book and expanding the material in the text, I'm interested in what they're saying and I'm learning something, but when they're just reading directly out of the book, I feel I can get this by reading out of the book myself.

- Do you think the discussion sessions affected your ability to understand the material?

> Because people that had questions that I thought I had known would bring up a point, and yes it did help to learn the material, because they would bring up a point I hadn't thought of or an angle I hadn't seen. Sometimes when I didn't understand, I could ask my question and because he encouraged discussion it would lead to something, rather than a quick yes or no answer.

At least one study [17] supports this anecdotal evidence that the conversational classroom works well for advanced courses, but what about classes dealing with cutting-edge research? Other techniques may be necessary when there is no satisfactory textbook or reference material.

In this paper, we have demonstrated that the conversational classroom strategy provides a sustained gain in academic performance. Even though many students complain, and attendance probably decreases, *overall the class does better*. That's the bottom line: using the classroom sessions in a conversational way just gives better results.

Acknowledgments

This work was supported by the National Science Foundation under grant EIA 008625.

7. REFERENCES

[1] Computers as components. http://ece.colorado.edu/~ecen2120.

[2] Gradekeeper. http://ece.colorado.edu/~ecen2120/gradekeeper/.

[3] M. L. Johnson Abercrombie. *The Anatomy of Judgement*. Hutchinson, London, 1960.

[4] P. G. Armour. The five orders of ignorance. *Communications of the ACM*, 43(10):17–20, 2000.

[5] Benjamin S. Bloom. *Taxonomy of educational objectives*. Allyn and Bacon, Boston, MA, 1984. http://www.coun.uvic.ca/learn/program/hndouts/bloom.html.

[6] Kenneth A. Bruffee. The art of collaborative learning: Making the most of knowledgeable peers. *Change*, 26(3):39–44, 1994.

[7] Louis L. Bucciarelli and Sarah Kuhn. Engineering education and engineering practice: Improving the fit. In Stephen R. Barley and Julian E. Orr, editors, *Between Craft and Science*, Technology and Work, pages 210–256. Cornell University Press, 1997.

[8] Computing Accreditation Commission. *Criteria for Accrediting Computing Programs*. ABET Inc., Baltimore, MD, 2004. http://www.abet.org/forms.shtml.

[9] Douglas Duncan. *Clickers in the Classroom*. Benjamin Cummings, 2005.

[10] Michele H. Jackson and April R. Trees. The learning environment of clicker classrooms. *Learning, Media, and Technology*, 2006. http://www-plan.cs.colorado.edu/pltools/pubs/F03clickers.pdf.

[11] Eric Mazur. *Peer instruction: a user's manual*. Prentice Hall, Upper Saddle River, NJ, 1997.

[12] Rebekah Nathan. *My Freshman Year*. Cornell University Press, Ithaca, NY, 2005.

[13] David. A. Schön. *The Reflective Practitioner: How Professionals Think in Action*. Basic Books, 1983.

[14] Sheila Tobias. *They're not dumb, they're different: stalking the second tier*. Research Corp., Tucson, AZ, 1990.

[15] William M. Waite, Michele H. Jackson, and Amer Diwan. The conversational classroom. In *Proceedings of the 34th ACM Technical Symposium on Computer Science Education*, pages 127–131, New York, 2003. ACM Press.

[16] William M. Waite, Michele H. Jackson, Amer Diwan, and Paul M. Leonardi. Student culture vs group work in computer science. In *Proceedings of the 35th ACM Technical Symposium on Computer Science Education*, pages 12–16, New York, 2004. ACM Press.

[17] Marilyn A. Wells and Phillip W. Brook. Conversational km — student driven learning. In Raymond Lister and Alison Young, editors, *Proceedings of the Sixth Australasian Computing Education Conference*, volume 30, pages 335–341, Sydney, 2004. Australian Computer Society.

[18] Julia T. Wood. Diversity in dialogue: Communication between friends. In Josina M. Makau and Ronald C. Arnett, editors, *Ethics of communication in an age of diversity*, pages 5–26. University of Illinois, Urbana, IL, 1997.