EVALUATION OF RESEARCH-BASED INNOVATIVE PEDAGOGY IN SEVERAL MATH COURSES

Mark Asiala, Ed Dubinsky

Report

During the three academic years from Fall, 1997 through Spring, 1999 a number of mathematics faculty and staff at Georgia State University have used certain innovative pedagogical strategies in several different courses. This study is an attempt to assess the effectiveness of our approach to teaching in terms of the goals of improving both math performance of students and students' attitudes towards mathematics.

We begin with an overview of the the literature on how others have tried to achieve the same goals at other universities. We then give our results based on data obtained from the registrar and a survey of students. This is followed by a discussion of what these results might mean.

Our results suggest that this pedagogical approach can be used at Georgia State without serious dificulty. There is some evidence that it leads to an improvement in student learning and that students who experience it are more likely to finish a course sequence and also tend to go on to take more courses in mathematics than do students experiencing traditional pedagogy. Finally, the students' attitudes towards this approach do not appear to be extreme in either a positive or negative direction while they are taking the course, and it improves with time.

1 Literature Review

There is no established method of evaluating the effectiveness of teaching in collegiate mathematics that has widespread acceptance. As Tucker and Leitzel (1995) argue, this is in part a reflection of the amount of autonomy given to individuals and departments in the design of their courses and is in part because the value of a college course is difficult to pin down. However the growth of calculus reform in colleges has highlighted the importance of evaluating teaching efforts and thus there is a body of literature that attempts to address the efficacy of innovative teaching. This review will outline the methods of evaluation that have been used to assess undergraduate mathematics courses in an attempt to provide some idea of the components that a comprehensive evaluation could entail. We will then look at how the design of the current research project relates to the existing body of literature.

Although there has been in recent years some work in reform of collegiate mathematics courses other than calculus, there are almost no attempts to evaluate its effectiveness. Therefore, most of this literature review is related to calculus.

The NSF Working Group on Assessment in Calculus (Schoenfeld, 1997) makes it clear that meaningful assessment of any course requires a careful exposition of the instructional goals of the course and a gathering and analysis of data that correspond to these. The evaluation of a course assesses whether the instruction in the course fosters the type of student learning and understanding it sets out to obtain. Such evaluation not only provides an indication of how successful a program of instruction is, but also can give us a deeper understanding of how students learn mathematics and what we can do to improve this learning. A variety of studies have been made along these lines. Some have taken the path of comparing students in a traditional calculus course with those in a reformed calculus course. Bookman and Friedman (1994) made comparisons between traditional and reformed calculus students in problem-solving performance where they found that students in the lab-based course (Project CALC) outperformed students taking traditional in problem solving. Park and Travers (1996) compared Calculus and Mathematica students with traditional students on achievement, attitudes and understanding. The findings favored the Calculus and Mathematica students. Of particular interest in this study is the use of concept maps as a method for probing conceptual understanding. Making concept maps requires students to externalize their thinking by mapping out the interrelationships between concepts and within concepts. Park and Travers argue that assessments of this nature which go beyond the usual comparison of test scores are important for assessing the effects of projects on students' understanding of calculus. Roddick's (1995) small-scale, but in-depth analysis of students' application of calculus knowledge to solve problems in an engineering mechanics course showed that students from a Calculus and Mathematica course were more likely to solve problems from a conceptual viewpoint than the traditional students who tended to focus on procedures. Meel (1998) also presents a small-scale, in-depth analysis comparing students from a Calculus and Mathematica course with students from a traditional calculus course on their understanding of key calculus concepts.

Of particular relevance to the evaluation of the C^4L calculus reform project, and in line with the goals outlined by the NSF working group on assessment, is the body of research evidence that underlies the C^4L program (See, for example, Asiala, 1996; Cottrill, et al., 1996; Clark, et al. 1997.) This research differs from the work mentioned above in that there is no attempt to directly compare traditional and C^4L students on their understanding of concepts. The research on C^4L students' understanding of various concepts forms part of coherent program of research based on a theory of learning. The research is of critical importance in any evaluation of the C^4L project. The goals of the research are "to increase our understanding of how learning mathematics takes place, to develop a theory-based pedagogy for use in undergraduate mathematics instruction, and to develop a base of information and assessment techniques which shed light on the epistemology and pedagogy associated with particular concepts" (Asiala et al., 1996, p6). The C^4L project is thus rooted in a research-based approach to curriculum development and a large part of the assessment of the project is bound up in the detailed investigation of student understanding that this body of research contains.

Although the effects of innovative courses on student understanding is key in the evaluation of such courses there is obviously also a need to address their effects on student learning on a broader and more global level. Comparing the attitudes towards mathematics, retention and continuation rates, grades and study habits of students in traditional courses with students in innovative courses is important to provide evidence of whether the innovative courses are successful in promoting student achievement and interest in mathematics. However, designing studies that allow one to make such comparisons meaningfully is difficult. Each of the factors is easily confounded. For example: Do better retention and pass rates reflect a better course or simply easier tests? How do we interpret self-reported attitudes? A challenging course might produce initial negative attitudes which may well change when the students begin to see the benefits of the work they've put in. Various researchers have dealt with such difficulties in different ways and we provide here a brief overview of the type of research that has been done and highlight those studies we have drawn on in the design of our study.

Measuring differences in student performance has been done either by comparing the grades of students in traditional and reform courses (Schwingendorf, submitted, Ratay, 1993) or by comparing the performance of students on a common examination/task (the Baylor and Naval comparison studies reported in Tucker and Leitzel, 1995, Bookman and Friedman, 1994, Meel, 1998, Park and Travers, 1996). In such comparison studies the researchers are often forced to make use of intact groups and most of the researchers point to the need to carefully examine these groups for differences. Similarly, in most cases, students self-select either the traditional or reform course which raises concerns of validity. Schwingendorf and McCabe use an analytic method to deal with these issues statistically.

Improvement in student attitudes is often cited by faculty to be one of the noticeable benefits in innovative courses (Tucker and Leitzel, 1995) and many of the studies evaluating innovative courses have included a component looking at attitudes. (Bookman and Friedman, 1998, Park and Travers, 1996). Keith (1995) describes students reaction to calculus reform and argues that we need to look at data about attitude and learning together in order to sensibly evaluate these. In Bookman and Friedman's evaluation of Project CALC they found that students tended to dislike the course during the first months, but that their attitudes change over time. Their survey of students one and two years after the course revealed significant differences between traditional and Project CALC students in terms of their beliefs about mathematics and attitudes towards mathematics which suggests that such longer-term studies of student attitudes are needed. Berenson, Vidakovic, and Haynie (1996), whose attitude test was adapted for our study, reported similar results. Again it is often reported that students in reform courses spend more time on the course and are more actively engaged in the work. (Tucker and Leitzel, 1995). Bookman and Friedman's evaluation of Project CALC provided evidence of this. Mathews (1995) showed that C^4L students generally appeared to be spending significantly more time studying calculus than the students in the comparison traditional calculus group. The Schwingendorf (submitted) study of C^4L students points to the possibility that the students see the rewards of this extra work in higher grades for calculus and with no accompanying adverse effects on grades for other courses. They suggest it is thus also important to compare the overall grades of students in order to ameliorate concerns that reformed courses will detract from student performance in other subject areas.

The current study is closely modeled on the Schwingendorf (submitted) longitudinal study comparing C^4L calculus students with traditional calculus students at Purdue University. The research on student learning and understanding on which the C^4L project rests provides a basis for evaluating the success of the project in that regard. Our study thus focuses on evaluating the success of the approach at Georgia State University in relation to student attitudes towards mathematics, retention and continuation rates, and performance as measured by grades.

One of the few reports evaluating curriculum reform in courses beyond calculus is by Asiala et al (1997) and concerns abstract algebra. In this study it is shown that the same pedagogical approach in the C^4L project and that studied here leads to improvements both in student learning and attitudes towards abstract mathematical concepts.

2 Description of Study

During the period covered by this study, a total of 11 sections of various courses were taught by two faculty, one part time instructor, and two graduate teaching assistants, using pedagogy based on research in learning. The pedagogical strategies included the use of computers to perform mathematical operations, writing programs to build tools that would perform mathematical operations, cooperative learning, and a de-emphasis of lecturing in favor of students working on tasks from worksheets.

We took as our population all students who took any version (experimental or standard) of these courses during the term indicated. For all analyses except the withdrawal rate analysis, we restricted our population to just those students who had completed the course. We removed all students who had received the grades, "AU", "AP", "CP", "I", "W", or "WF".

2.1 Courses

The courses involved were precalculus (one term), calculus 1 (three terms), calculus 2 (three terms), discrete mathematics (two terms) and abstract algebra (two terms). From the registrar we obtained data for the students who took these courses both in sections using our methods (experimental) and in sections using standard pedagogy (standard). The data included withdrawal rates, grades and courses taken. In addition we have used the course evaluations from the experimental courses and student responses to a questionnaire sent to to a random selection of students who took these courses.

The courses and terms were:

1. MATH126 FALL 1997 (Pre-Calculus)
2. MATH211 WINTER 1997 (Calculus I)
3. MATH211 WINTER 1998 (Calculus I)
4. MATH2211 FALL 1998 (Calculus I)
5. MATH212 SPRING 1997 (Calculus II)
6. MATH212 SPRING 1998 (Calculus II)

7. MATH2212 SPRING 1999 (Calculus II)
8. MATH220 FALL 1996 (Discrete Mathematics)
9. MATH2420 FALL 1998 (Discrete Mathematics)
10. MATH441/641 WINTER 1998 (Abstract Algebra)
11. MATH4441/6441 SPRING 1999 (Abstract Algebra)

2.2 Survey

The total population of students was estimated to consist of 1800 individuals from 75 sections of precalculus, calculus, discrete mathematics, and abstract algebra. Most of the population had already completed their course and the people still taking it at the time were approximately 2/3 of the way through the course.

Since it was not feasible to send out a questionaire to our entire population, a random sample of 400 individuals were selected and stratified by section. We planned for a minimum of 5 people per section and used the 25 surplus individuals for the largest sections. Within each section we ordered the individuals by their student identification numbers and then performed a systematic sample using a large enough skip number so that we were confident to traverse the entire roll for even the largest section.

We sent out 400 questionnaires with instructions to return it in the postage-paid envelope provided within 10 days. We received back over the course of 30 days 90 responses for a 23% response rate. The responses were nearly evenly divided between our experimental group and the traditional group. The response rate was slightly higher for the experimental group which is not suprising since they may have felt more in touch with the questions concerning group work and using technology than the traditional group. However, the difference was not large. The students were asked to respond to the following statements on a scale of 1 (Strongly disagree) to 5 (Strongly agree).

Question #1 "I enjoy doing math problems."

Question #2 "To solve math problems you have to know the exact procedure for each problem."

Question #3 "In mathematics, an answer is either right or wrong."

Question #4 "Mathematics requires much more memorization than understanding."

Question #5 "In the long run, I think mathematics will help me."

Question #6 "In order to understand math, I need to know more theory."

Question #7 "I learn math through examples."

Question #8 "Guessing is OK to use in solving mathematics problems."

Question #9 "My calculus course seems to require more thinking than memorization."

Question #10 "I think that I can apply what I'm learning in calculus in some other courses."

Question #11 "The best way to learn calculus is to memorize all the formulas."

Question #12 "Some people are good in calculus, some just aren't."

Question #13 "I like the idea of using technology in calculus."

Question #14 "The math teacher is responsible for how much math I learn."

Question #15 "A good math teacher demonstrates procedures for doing the type of problems that'll be on the test."

Question #16 "Good math teachers show students lots of different ways to look at the same question."

Question #17 "I like the idea of working in groups in a calculus course."

Question #18 "I feel uncomfortable when I see that other students know more math than I do."

In addition, students were asked (Question #19) to list three or more most valuable things they learned in their calculus course(s) and (Question #20) to list three or more most useless things they learned in their calculus course(s). Finally, the students were invited to make general comments.

Some bias is inevitable when dealing with questionnaires sent out in this manner but since this bias would be expected to appear in both groups, it should not be very strong. We would expect those with the strongest opinions concerning their mathematics experiences to respond. If that bias was different in the two groups, we might expect the variance on a question for one group to be significantly different from the other group. However, we find that the limitation on the variance had more to do with being at one end of the response spectrum (near 4 or 5) or the other (near 1 or 2) than any true significant difference between the two. Overall, in the few cases where the variance was significantly different, it only affected the p-value in the third or fourth decimal place.

3 Results

Our results consist of comparisons of the experimental sections with corresponding standard sections of the same courses taught at the same time. We compare withdrawal rates, student grades, the number of math courses taken for which the courses we are studying are prerequisites, grades in these subsequent courses, summaries of the student evaluations, and of the questionnaire responses.

For the comparisons, we omitted the abstract algebra sections because no standard section was taught at the same time. We also omitted the latest calculus 2 section because it was too recent to obtain data on subsequent courses taken. Finally, there are no comparisons of the student evaluations in the experimental courses with those from the traditional courses because to organize the latter data is beyond the resources of this study.

In addition, the statistical analysis factored out the affect of differences in student abilities by using the predicted grade point aveerage calculated by the registrar for every student.

3.1 Withdrawal rates

The withrawal rates in the experimental and traditional courses were virtually identical. The actual difference was less than 1% and this difference was not statistically significant ($p > 0.1$).

3.2 Grades in courses

In all 8 sections the students who took the experimental courses received higher grades than did those taking the standard courses. The grades (measured as grade point average with 4 for "A", 3 for "B", etc.) ranged from 1.81 to 2.86 and the differences ranged from 0.05 to 0.94. In 4 of the courses the differences were statistically significant with $p < 0.015$

3.3 Subsequent math courses taken

From 6 of the 8 sections, the students who took the experimental section took more math courses for which this course was a prerequisite than did the students who took the standard section. The average number of subsequent courses taken ranged from 0.11 to 2.38 and the differences ranged from 0.02 to 1.33. The difference was statistically significant ($p = 0.005$) in one case.

For the other two sections, the average number of subsequent courses taken ranged from 0.11 to 1.16 and the differences were 0.01 and 0.21. Neither difference was statistically significant ($p = 0.09$).

One factor affecting how many courses a student took is the timing. Obviously, students near the end of their program will take less courses than those just starting. The effect of this factor was not taken into account in our analysis except that we can expect that the time a student takes a course will be independent of whether it is taken in an experimental or standard section.

3.4 Grades in subsequent math courses

From 6 of the 8 sections, the students from the standard courses received higher grades in the subsequent math courses than did the the students from the experimental courses. The average grades ranged from 1.85 to 2.77 and the differences ranged from 0.05 to 0.65. In only one case was the difference statistically significant ($p = 0$).

For the other two sections, the students from the experimental courses did better. The average grades ranged from 1.44 to 2.64 and the differences were 0.2 and 0.42. Neither difference was statistically significant ($p = 0.9$).

3.5 Student evaluations of the experimental courses

The data reported in Figure 1 is for all of the 11 experimental courses totaling 25 sections. The evaluation is the students' responses to the question: Considering both the limitations and possibilities of the subject matter and course, how would you rate the overall teaching effectiveness of this instructor? The responses are scored on a scale of 1-5 with the higher score indicating the greater effectiveness. The number given is the average over all students in all experimental sections of the course in a given term.

If we take an overview of the five courses: precalculus, calculus 1, calculus 2, discrete mathematics and abstract algebra, it appears that with one exception, the averages cluster around the middle. The exception is calculus 2, in which many, but not all students were taking the second semester ofthe experimental course. Here the score is somewhat higher.

Course	# Sections	Evaluation
Ma126 Fall 97	3	3.3
Ma211 Winter 97	1	2.9
Ma211 Winter 98	4	4.1
Ma2211 Fall 98	5	3.2
Ma212 Spring 97	1	4.1
Ma212 Spring 98	1	4.2
Ma2212 Spring 99	4	4.0
Ma220 Fall 96	2	2.8
Ma2420 Fall 98	2	3.3
Ma441 Winter 98	1	3.3
Ma4441 Spring 99	1	2.6

Figure 1: Student Evaluations in Experimental Courses

3.6 Attitudes about math

The student responses to the questionnaire appear to tell us something about the attitudes they develop, to some extent towards the end of the course (and thus should be closely related to their evaluations) but mainly at some time after the course is completed and they have other, perhaps related, mathematical experiences.

Following is a tabulation of the responses on the 18 questions. Recall that the responses were on a 1-5 basis with 5 indicating strong agreement. The p-value is for the one-tail t-test[1]. In the Desired Outcome column we indicate whether a positive response or a negative response is more consistent with the goals of the course. Thus, we certainly want the students to enjoy doing math, so a positive response is desired for the first question. On the other hand, we want students to accept the fact that it is necessary to look for methods other than those identical to detailed procedures they already know so a negative response is preferred on the second question.

On 16 of the 18 questions, the responses of the students from the experimental sections differed from the students from the traditional section. In 11 of these, the difference was more than 10%, and in 7 of these, the difference was statistically significant at the 95% confidence level.

For the two cases in which the direction of difference favored the traditional courses, in one case the difference was less that 0.1% and in the other case it was more than 10%.

[1]On Questions #7, #14, this test was considered inappropriate.

	CL		TRAD		Desired Outcome	p-value
Q#	Mean	Var	Mean	Var		
1	4.09	1.06	3.72	0.92	+	0.039
2	3.07	1.51	3.17	1.35	−	0.338
3	3.07	1.73	3.54	1.36	−	0.038
4	1.70	0.68	2.15	1.02	−	0.012
5	4.39	0.75	4.28	0.56	+	0.273
6	3.33	1.18	3.24	1.12	+	0.352
7	4.55	0.44	4.51	0.26	−	N/A
8	2.67	1.27	2.35	0.90	+	0.071
9	3.86	0.88	3.39	1.18	+	0.016
10	3.76	1.09	3.42	1.16	+	0.065
11	2.20	0.86	2.57	1.01	−	0.040
12	2.77	1.30	3.17	1.08	−	0.042
13	3.86	1.19	3.59	0.96	+	0.104
14	3.18	1.50	2.83	1.39	−	N/A
15	3.93	1.23	4.13	0.74	−	0.173
16	4.55	0.30	4.14	0.67	+	0.004
17	3.57	1.79	3.22	1.69	+	0.105
18	2.68	1.66	3.00	1.24	−	0.107

Figure 2: Means and Variances of Questionnaire Responses

4 Discussion

Although all aspects of the experimental courses are clearly different for the students, and there are some anecdotal reports of their dissatisfaction with relying on members of their group, working on computers and being asked to figure out many mathematical issues on their own (with help from the instructor), there were no serious incidents of formal complaints to the administration, and the students were at least as likely to complete the course as were the students from the traditional sections. Everything that we know is consistent with the suggestion that the reports of difficulties have mainly to do with the natural rough edges created by any pedagogical approach that is highly unfamiliar to students and to other faculty members.

4.1 Student learning

The fact that the students received substantially higher grades in the experimental sections than in the traditional sections could indicate that they were more successful in learning the material. It is also possible, however, that there was some kind of grade inflation effect in the experimental sections.

It is the case that students worked in cooperative groups and some of their grades were affected by how much and how well their fellow group members learned. Of course this could work in two ways and on the average one might expect excesses in both directions to be muted. Moreover, a substantial amount of the students' grades were based on performance that was entirely individual. For the most part, it turns out that final grades of the students were not very different from their scores on individudal exams and quizes. Data is available to confirm this if an appropriate effort to organize it were to be made.

That leaves the possibility that the student assessment instruments were easier in the experimental sections than in the traditional sections. Again, there is reason to believe that this is not the case. Exams in the experimental sections often asked students to solve problems they had not been taught how to solve and essay questions were effective in elucidating understanding rather than rote repetition of learned procedures. Anecdotal reactions from students suggest that the exams in the experimental sections were actually harder. All of the material should be available and again an appropriate effort could quantify this point.

The greater tendency of students from the experimental sections to go on in their studies of mathematics seems to be an unequivocal result and is consistent with what is found in other studies. Regardless of how students may feel at the time of taking the course, the higher grades (which we feel are justified) and the more positive feelings about mathematics (see the discussion of the survey results below) seems to make the students less anxious about mathematics and more willing to pursue it further.

On the other hand, the students from the experimental sections do not do quite as well in these subsequent courses (almost invariably using traditional pedagogy) as do the students from the traditional sections. Although the differences are small (less than half a letter grade on average) and with one exception, not statistically significant, this is a matter of concern and is one area that implies the need for further thought about ways to improve our approach. The first step in such thinking is to try to find possible reasons why the students from the experimental courses did not do better in subsequent mathematics courses.

We can think of at least three possible reasons. First, it could be that all of the other indications are illusory and the students taking the experimental sections simply have not learned the material any better than the students in the traditional sections. A second possible reason is based on the suggestion that all of the students who normally go on to take more mathematics courses are pretty good in this subject. Therefore, the effect of the experimental sections is to get more students not quite as strong as others to take more mathematics courses. Although this would certainly be a desirable effect, it could explain the fact that these somewhat weaker students do not excel in the higher level courses. Finally, the phenomenon might be explained by the differences in the nature of the experimental and traditional approaches. It could be argued that the latter approach focuses almost entirely on rote learning and more or less gives up on conceptual understanding, whereas the experimental approach pays less attention (although still substantial) to learning procedures and more on conceptual understanding. If these students go on to courses which expect them to have been highly trained in their previous courses to follow procedures they are shown and repeat them on examinations and do not give them an opportunity to demonstrate what may be a greater conceptual understanding, then it would not surprising that they do not perform any better.

Nothing in the information we can present in this report helps us decide among these three alternatives and further study would be required to resolve the question.

4.2 Student attitudes

The scores on the student evaluations are reasonable. In four of the 11 cases they are 4 (out of 5) or higher and in three cases they are a little below 3. It is interesting to note that in the one course in which many students are experiencing the experimental pedagogy for the second time, all three courses average at least 4. There are two possible explanations for this. One is that only the students for whom this approach works stay with it for the second course, the others transfer to sections using traditional pedagogy. The other explanation is that as students get more used to this method, their attitude becomes more positive.

The results of the questionnaire, which was taken well after the students had finished the course is consistent with the idea that student attitudes about this pedagogical approach improve with time.

We consider the various questions on the questionaire in categories.

Questions #13, #17 are the same kind of approval issues that we have on the student evaluations. For the students in standard sections we are asking about the idea of using technology whereas for the experimental sections it is based on their actual experiences. A closer look at the high numbers from both groups shows that the higher scores from the experimental sections come mainly from more students indicating strong agreement. This suggests that students from both groups like the idea of using technology and cooperative learning and that the actual experience with these pedagogcial strategies tended to confirm that view.

Essentially the same can be said about Questions #1, #5, #10 and these questions represent important attitudes about mathematics. It is encouraging that all of the students seem to have positive attitudes and that their experiences in the experimental sections appeared to have strengthed these feelings.

With Questions #12, #18 we move onto the all important issue of student confidence in their approach to mathematics. These are responses that can be taken at face value (if someone says they feel confident, then they probably do) and what may indicate an improvement on the part of the students from the experimental sections is a very good sign, although the scores from both groups are higher than we would prefer.

This is related to Questions #14, #15, #16 which are asking about a students' willingness to rely more on her or himself than on the teacher for learning mathematics. It is disappointing that the students from the experimental sections seem to want to rely more on the teacher than do the students from the traditional sections, but it is encouraging that in the other two questions they indicate more of an understanding of what a teacher must do in a situation where the student takes on more responsibility for learning. In the case of Question #14, where the students from the standard sections disagreed more strongly, the written comments suggested that their responses were based on their difficulties with the teacher (poor English, illegible handwriting) and therefore a self-reliance in order to learn in spite of the teacher. For Question #16, the very high scores from the students in the experimental section could have been due to the fact that in these sections there was extensive use of the practice of carrying single examples throughout the course to illustrate a variety of ideas. It is possible that these scores represent an approval of that approach.

Questions #2, #3, #4, #8, #11 represent important attitudes about the nature of mathematics that are not held by students in general. The survey results on these questions in both the amounts of differences with students in the traditional courses and the consistency of those differences suggest that the experiences in the experimental sections may have led to important and desireable changes in student attitudes about mathematics. In addition to that, the responses on Questions #6, #9 suggest a greater acceptance of the role of theory in mathematics on the part of the students in the exprimental sections. On the other hand, although these scores represent some progress they are much closer to the middle than we would prefer so that more progress is still needed.

Finally, the response on Question #7 suggests that the experimental sections did not contribute much to convincing the students that mathematics is not just about organizing information to be found in a host of examples, but also involves the study of structures that produce these examples. The primacy of examples appears to be a very deep belief reinforced by many years of experience in courses which focus on examples as opposed to structures. It is possible that one reason for the high agreement with this statement by students from the experimental sections is based on the use of problem solving which may appear to students as a focus on examples although it is a different use of examples than in the traditional sections.

5 References

Asiala, M., Brown, A. , Devries, D., Dubinsky, E., Mathews, D. and Thomas, K., (1996). A Framework for Research and Curriculum Development in Undergraduate Mathematics Education. CBMS Volume 6 p1-p32

Asiala, M., Brown, A., Clark, J., DeVries, D., Dubinsky, E., Hemenway, C., Mathews, D., Morics, S., Okta, A., St. John, D., Thomas, K., Tolias, G., and Vakil, R., (1997). An investigation of Students' Understanding of Abstract Algebra and the Use of Abstract Structures to Build Other Structures in E. Dubinsky (ed.) Journal of Bathematical Behavior, 16,3, pp. 1-309.

Berenson, S., Vidakovic, D., & Haynie, G. (1996). Laboratory Improvement for the Calculus Curriculum. Final Report to the NSF Principal Investigator. Raleigh, NC: Center for Research in Mathematics and Science Education.

Bookman, J. and Friedman, C. (1998) Student Attitudes and Calculus Reform. School Science and Mathematics, March 1998: p117-p122 Bookman, J. and Friedman, C. Evaluation of Project CALC (Website address)

Bookman, J. and Friedman, C.P. (1994) A Comparison of the Problem Solving Performance of Students in Lab Based and Traditional Calculus. CBMS Volume 4 p101-p116.

Clark, J., Cordero, F., Cottrill, J., Czarnocha, B., DeVries, D., St. John, D., Tolias, G., & Vidakovic, D. (1997). Constructing a schema: The case of the chain rule. Journal of Mathematical Behavior, 16(3).

Cotrill, J., Dubinsky, E., Nichols, D., Schwingendorf, K., Thomas, K., & Vidakovic, D. (1996). Understanding the limit concept: Beginning with a coordinated process schema. Journal for Mathematical Behavior, Volume 15, pp. 167–192.

Keith, S.Z. (1995) How Do Students Feel about Calculus Reform and How Can We Tell? UME Trends v6 n6 p6, 31.

Mathews, D.M. (1995) Time to Study: The C4L Experience, UME Trends v7 n4 Meel, D.E. (1998) Honors Students' Calculus Understandings: Comparing Calculus & Mathematica and Traditional Calculus Students. CBMS Volume 7 p163-p199

Ratay, G. (1993) Student Performance with Calculus Reform at the United States Merchant Academy. Primus v3 n1 p107-p111

Roddick, C.S. (1995) How Students Use Their Knowledge of Calculus in an Engineering Mechanics Course. Paper presented at the 17th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Columbus, OH.

Schoenfeld, A. (ed) (1997) Student Assessment in Calculus. Washington, DC: Mathematical Association of America

Schwingendorf, K.E., McCabe G. E. & Kuhn, J. (submitted) A Longitudinal Study of the C^4L Calculus Reform Program: Comparison of C^4L and Traditional Students.

Travers, K. J. and Park, K. (1996) A Comparative Study of a Computer-Based and Standard First-Year Calculus Course. CBMS Volume 6 p.155-p176

Tucker, A.C. and Leitzel, J.R.C. (1995) Assessing Calculus Reform Efforts. Washington, DC: Mathematical Association of America