# A Longitudinal Study of the C$^4$L Calculus Reform Program: Comparisons of C$^4$L and Traditional Students

Keith Schwingendorf, George McCabe, and Jonathan Kuhn

*The authors are extremely grateful to each of the following people for their support, helpful discussions, comments and/or insightful suggestions for improving this paper: Linda McCabe who served as a statistical consultant and without whom this paper could not have been completed; Ed Dubinsky for his ideas regarding the analyses of the data; David M. Mathews, Jim Cottrill (Illinois State University), Lisa Schwingendorf, and Shonda Kuiper for their careful reading of the manuscript and suggestions for improvements of the paper. A special thanks goes to the 205 Purdue students reported on in this paper for their patience, hard work and dedication to the goals and principles of the C$^4$L Calculus Program.*

Abstract. This paper presents the results of a longitudinal statistical study in which comparisons are made between students who were taught introductory calculus courses using the Calculus, Concepts, Computers and Cooperative Learning (C$^4$L) pedagogical methodology and students taught in the traditional way (TRAD). Two basic, related, questions considered in this study are:

- Which program, C$^4$L or TRAD, provides a student with a better understanding of the required calculus concepts?
- Which program, C$^4$L or TRAD, better inspires students to pursue further study in calculus or, more generally, mathematics?

The results of this study favor the C$^4$L program over the TRAD program. On the one hand, for example, it is found the C$^4$L students earn higher grades in calculus courses; in fact, almost half a grade higher, on average, than the TRAD students. On the other hand, for example, it is found the C$^4$L students are as prepared as the TRAD students, but not more so, for mathematics courses beyond the calculus program. After some discussion of the C$^4$L program, the paper describes the statistical model used to perform the comparison between the two teaching methods and then presents the results of this statistical analysis. This paper reports on one phase of the evaluation of the C$^4$L calculus program, a quantitative evaluation based on a statistical model which is similar to that used in some health studies (e.g., studies on the relationship between second-hand smoke and cancer) and industrial studies (e.g., studies of the effects of the notion of Total Quality Management as related to productivity and employee satisfaction) which are observational in nature. The other phase of the C$^4$L evaluation involves qualitative research studies involving both C$^4$L and TRAD students.

This research has been partially supported by the National Science Foundation grants # USE-9053432 and # DUE-9450750.

## 1. The C⁴L Program

The Calculus, Concepts, Computers and Cooperative Learning Calculus Reform Program is part of the National Calculus Reform Movement which began at the Tulane Conference (sponsored by the Sloan Foundation) in January 1986. The $C^4L$ program is currently co-directed by Dubinsky, Schwingendorf, and David M. Mathews (Southwestern Michigan College, Dowagiac, MI). Numerous papers and reports have been written or presented at national conferences in connection with the $C^4L$ program ([**2**], [**3**], [**4**], [**9**], [**6**], [**8**], [**7**], [**10**], [**16**], [**12**], [**24**], [**22**], [**23**], [**26**], [**31**], [**32**], [**34**], [**33**], [**30**], [**35**]). The $C^4L$ program texts ([**18**], [**17**]) were written in an effort to support the $C^4L$ pedagogical approach which is based on a constructivist perspective on how mathematics is learned ([**2**], [**3**], [**4**], [**9**], [**6**], [**8**], [**16**], [**12**], [**18**], [**17**], [**23**], [**33**]). According to the Action-Process-Object-Schema (APOS) theory of learning theory ([**2**]) on which the program courses are based, students need to construct their own understanding of each mathematical concept.

$C^4L$ calculus courses differ from traditionally taught calculus courses and also from courses in other calculus reform programs in a number of ways. The traditional courses, delivered primarily via the lecture and recitation system, in general, attempt to "transfer" knowledge, emphasize rote skill and drill together with memorization of paper-and-pencil skills. In contrast, the primary emphasis of the $C^4L$ program is to minimize lecturing, explaining, or otherwise attempting to "transfer" mathematical knowledge, but rather to create situations for students to foster their making the necessary mental constructions to learn mathematics concepts. The emphasis of the $C^4L$ program is to develop student understanding of concepts together with the acquisition of necessary basic skills through active student involvement in what is termed the Activity-Class-Exercise (ACE) learning cycle.

Each unit of the ACE learning cycle, which generally lasts about a week, begins with students performing group computer activities in a laboratory setting in an effort to help students construct their own meaning of mathematical concepts and reflect on their experiences with their peers in a cooperative learning environment. Students use a mathematical programming language (MPL), such as *ISETL*, or that which is part of the *MapleV* system ([**16**], [**15**], [**12**], [**13**], [**11**], [**24**], [**31**], [**33**]) to write "programs" which model mathematical concepts, then investigate the concepts using various computer tools. Such tools include graphics and table facilities to help students reflect on their experiences in an effort to foster mental constructions of mathematical concepts. Many computer activities have students transfer information, sometimes requiring a reconstruction, back and forth between a formula in a symbolic computer system (SCS) such as *Derive*, *Mathematica*, or *MapleV*, the constructions in the MPL, and graphs or tables constructed on the computer. Laboratory periods are followed by class meetings during which carefully constructed "class tasks" and appropriate questions in conjunction with cooperative problem solving are used to help students to build upon their mathematical experiences in the computer laboratory. Finally, relatively traditional exercises are assigned to reinforce the knowledge students are expected to have constructed during the first two phases of the learning cycle.

The study reported on in this paper is one of several studies and reports that are beginning to emerge regarding the long-term effects of Calculus reform ([**5**], [**21**], [**20**], [**19**], [**29**]). By long term, we mean studies done at least two to four years after students have completed their calculus courses. Susan L. Ganter reports ([**20**], [**19**]) the following trends concerning student performance:

> Evaluations conducted as part of the curriculum development projects revealed better conceptual understanding, higher retention rates, higher scores on common final exams, and higher confidence levels and involvement in mathematics for students in reform courses versus those in traditional courses; the effect on computational skills is uncertain.
>
> ... in general, regardless of the reform method used, the attitudes of students and faculty seem to be negative in the first year of implementation, with steady improvement in subsequent years if continuous revisions are made based on feedback.

In her conclusions, Ganter also states that

> The success or failure of a reform effort is not necessarily dependent upon what is implemented but rather how, by whom, and in what setting." This raises important issues of design and interpretation of evaluation studies and points to just one of many factors that can complicate matters.

The C$^4$L program qualitative research studies ([**3**], [**4**], [**9**], [**6**], [**8**], [**7**], [**23**]), which included both C$^4$L and TRAD students, generally suggest that C$^4$L students have a deeper understanding of the conceptual nature of the calculus and computational skills as good as those of their TRAD counterparts (which was also suggested by the results on common final exams in 1988 and 1989). In addition, a recent study on the results of a questionnaire on "study time" by calculus students that was completed by 38 faculty who have taught using the C$^4$L program ([**22**]) indicates a positive side-effect: C$^4$L students' weekly study hours had mean, 9.37; sd = 4.27 while TRAD students' weekly study hours had mean, 5.36; sd = 2.41. Hence, C$^4$L students generally appear to be spending more time studying calculus than do their TRAD counterparts which suggests that C$^4$L students may reap the rewards for studying more than TRAD students in that they often receive higher grades in calculus. Moreover, C$^4$L students appear not to be adversely affected in their other courses, as had been conjectured by advisors, students and other faculty, because no statistically significant differences in C$^4$L students' grade point averages as compared to TRAD students were found in this study.

## 2. Statistical Method

We now describe the statistical model used to perform the comparison between the two teaching methods. We first identify the population for this study. Then, we state the specific questions asked that address the issues of whether or not students do better in and/or take more subsequent courses that require calculus for either the C$^4$L and TRAD programs. Next, we look at the problem of nonrandom selection (or self-selection) of students which occurs in this study and how we dealt with this problem. We note that the problem of self-selection is common to other studies like this one as has been noted by other researchers, in particular, by Jack Bookman in his evaluations of the Duke University PROJECT CALC ([**5**]). Lastly, we summarize briefly the primary statistical model used.

**Population.** The students in this study were those enrolled at Purdue University, West Lafayette, IN from Fall 1988 through Spring 1991 in either the traditional lecture and recitation three semester calculus sequence consisting primarily of engineering, mathematics and science students (TRAD); or the corresponding $C^4L$ reform calculus courses ($C^4L$). By "other" math courses we mean mathematics courses beyond the basic three-semester calculus sequence.

We made several comparisons of the 205 $C^4L$ students with 4431 TRAD students who took their first year of calculus in the academic years 1988, 1989 or 1990. Thus, a total of 4636 students were included in these analyses.

Only data on those students enrolled for the first time in: (1) beginning calculus during the fall semesters of 1988, 1989, and 1990; (2) second-semester calculus during the spring semesters 1989, 1990, and 1991; and (3) third-semester calculus during the fall semester of 1990 and 1991 were included in the study reported on in this paper.

The sample of data in the study might be considered, in one sense, the *population* of all data since it is the complete set of data for all students who took the indicated calculus courses over the given three year period at Purdue. Of course, we would like the results of this study to be used to infer conclusions about *all* students who *might* take the indicated calculus courses at Purdue University or, for that matter, more generally, a "typical U.S. University," say. To the extent that the 4636 students in the study are somewhat representative of these more general populations, the results in this study can be extended to these more general populations.

**Comparisons Made.** We now consider, the first, more important, question related to which program, $C^4L$ or TRAD, provides a student with a better understanding of the required calculus concepts. One possible way of deciding this would be to compare, say, the average final grades for the two programs. Because the final grade reflects a student's knowledge and understanding of the course material, the program whose students obtained the highest final average grade would be deemed the better program.

Three statistics were used in the present study to compare a student's level of understanding of calculus concepts under the $C^4L$ program with a student's level of understanding under the TRAD program and are given below.

- Average final grade.
- Average final grade for all mathematics courses taken after the completion of the calculus programs.
- Last available overall grade point average.

Instead of the average final *grade*, which includes both term component and a final mark component, it might have made more sense to compare the average *final grade* alone. Although common finals for first year calculus were given in 1988 and 1989 for the $C^4L$ and TRAD programs, common finals were not given in 1990, and so a comparison using final marks was not feasible. We should note that the common finals given in 1988 and 1989 were those written by the TRAD faculty, and the $C^4L$ average final exam scores were either about the same or higher than the TRAD averages ([**16**]). Common finals were not given in 1990 primarily due to a lack of agreement between TRAD faculty and the $C^4L$ faculty on the nature and scope of questions that should be included on the final exams.

Comparing the average final grade for all mathematics courses taken by students after the completion of calculus in the two programs shed some light on whether students performed better in other mathematics courses after having taken the $C^4L$ program or after having taken the TRAD program. Presumably, a student with a good understanding of calculus would perform better in a mathematics course that required (or did not require, for that matter) a knowledge of calculus than a student who had a poor understanding of calculus. Similarly, comparing the last available overall grade point average taken by students in the two programs gave some indication as to whether students performed better in *general*, for all courses, after having taken the $C^4L$ program or after having taken the TRAD program. Moreover, the demands on $C^4L$ students due to an additional workload (e.g., the laboratory component of the course) and the notion that this might interfere with $C^4L$ students' performance in other course was a consideration we wished to try to measure using this third statistic.

There are other ways of deciding which of the two programs provides a student with a better understanding of the calculus concepts. Certainly, one possible way of measuring this would be to attempt to conduct extensive impartial qualitative interviews with students from each course and, on this basis, decide which of the two programs provided a student with a better understanding of the calculus concepts. This has been done in several studies involving both TRAD and $C^4L$ students reported on in this paper ([**3**], [**4**], [**9**], [**6**], [**8**], [**7**], [**23**]).

Consider, now, the second basic question considered in this study, related to which program, $C^4L$ or TRAD, better inspires students to pursue further study in calculus and, more generally, mathematics. In this case, the interest a student showed in either calculus or mathematics, in general, was measured by counting the number of either calculus or mathematics courses this student took while pursuing his/her degree.

Two statistics considered in the study to compare a student's level of interest in calculus or mathematics courses under the $C^4L$ program with a student's level of interest under the TRAD program and are given below.

- Average number of calculus courses.
- Average number of non-calculus mathematics courses.

**The Issue of Non-Random Selection of Students.** Some students were encouraged to take the $C^4L$ reform calculus course by academic advisors in the School of Science at Purdue. Academic advisors informed students about the nature of the $C^4L$ course through a brief course description written by the original $C^4L$ program directors. Every effort was made to enroll more female students than is usual in the first-semester $C^4L$ Calculus course (MA 161A) and also to enroll as many prospective mathematics education teachers as possible (to provide a "role model" for teaching mathematics). The Director of the Women in Engineering program advised both female and male Freshman Engineering students and, in addition, coordinated and worked closely with other Freshman Engineering advisors in selecting students for the $C^4L$ courses from 1989 − 1991. Students were allowed to switch between the $C^4L$ and TRAD calculus courses if they so desired. (However, few students made the switch from MA 161A to the traditional MA 161.) Hence, in effect, students either were directed or self-selected themselves, or some combination of the two, into the $C^4L$ calculus courses.

Random assignment of students to the $C^4L$ or TRAD programs would have been preferable to the nonrandom approach taken in the study. Random assignment would have offset any possible confounding factors which might have influenced the conclusions of the present study. Two factors which probably did influence the study and which were not taken into consideration were the grading policies and the amount of time spent by the students in the $C^4L$ and TRAD programs.

The $C^4L$ students were strongly encouraged to work in their permanent assigned groups (of three or four students) throughout the course whereas TRAD students worked as individuals. Four of the six course grade items for $C^4L$ students were based on group achievement (where each individual was assigned the group grade except in the case when the individual did not fully participate) while two course grade items were based on individual achievement. All five course grade items for TRAD students were based on individual achievement. The $C^4L$ students were assigned grades according to a predetermined benchmark or standard point total while TRAD students were assigned grades according to the rank of their point total. For example, $C^4L$ students were each assigned a grade of A on a group test if the group score was 90% or higher, say, while TRAD students were assigned a grade of A if their point total on an exam was in the top 20 percent. In general, benchmark-grading $C^4L$ instructors were allowed greater flexibility and responsibility than rank-grading TRAD instructors in assigning course grades ([**14**] provides a more complete discussion of the $C^4L$ grading policy).

The $C^4L$ students spent seven in-class hours per week on course material (four in lab and three in class), whereas TRAD students spent five in-class hours on course material. The $C^4L$ students engaged in a variety of thought-provoking group activities in addition to participating in class discussions and summaries of class concepts. The TRAD students attended three lectures and two recitation classes.

The issue of nonrandom assignment of students to programs is a limitation of the present study. However, it would seem there are no practical alternatives to this design. Random selection would almost certainly imply assigning at least some of these students, against their wishes, to either the $C^4L$ program or the TRAD program. For most, if not all students, how well they perform academically is an overriding concern and so to place them in a course in which they have doubts about could possibly bias the results of the experiment. To randomly assign a student into one or the other programs without this student's consent raises ethical issues. Furthermore, an approach which did not take into consideration the informed consent of each $C^4L$ student would have ultimately lead to bias anyway because the two programs are so obviously different, a student would have quickly become aware of these differences.

**Dealing With Non-Random Assignment.** A possible solution to the problem of the nonrandom assignment of students to either the $C^4L$ or TRAD programs, is to "chip away at it" by making separate comparisons between smaller more homogeneous groups, defined by known extraneous factors, of $C^4L$ students and TRAD students, or, in other words, to *control* for known extraneous factors. We note that Bookman and other researchers also use the notion of control variables to deal with nonrandom assignment. For example, say academic ability is identified as a possible confounding factor with academic performance of $C^4L$ students and TRAD students. That is, suppose it is felt students are performing better academically not because they are taking the $C^4L$ program but simply because they are better students. In this case, the effect of academic ability on $C^4L$ students should be investigated and then, separately, the effect of academic ability on TRAD students would be investigated. Then, if there was shown to be a large positive association between academic ability and C$^4$L students and only a moderate positive association between academic ability and the TRAD students, say, this would indicate the students were performing better academically because of the C$^4$L program and not simply because they were better students.

Making comparisons between smaller more homogeneous groups of the C$^4$L or TRAD programs, to control for known extraneous factors, is not as effective as assigning students to programs at random. There could well be extraneous factors the authors were unaware of which are confounded with the type of teaching method program and so would not be dealt with by this procedure. Nonetheless, it appeared as though (at least) the following extraneous factors needed to be addressed as they appeared to be potentially confounding to the results of the study.

- Predicted grade point average (see below).
- Major course of study (the last available major).
- Gender.

A student's academic ability was measured using a modified grade point average called a *predicted grade point average*, or PGPA. This is a statistic which the registrar's office computes for entering freshman at Purdue to predict students' first semester grade point averages ([**25**], [**36**]). Using the 1971 beginning first year students as the population (1970 and 1971 combined for those schools within Purdue with small enrollments), multiple linear regression analysis was performed to determine which combination of available predictor variables were most strongly associated with the obtained first semester index. The best predictors were found to be SAT-Verbal, SAT-Math, average high school grade (a weighted average of the number of semesters taken and the grades received for math, English, science, modern language and speech), and high school percentile rank cubed (cubed rank was used since it spreads the distribution and reduces the skew). Prediction equations involving these four variables were developed for each of the schools within Purdue, and in some cases, specific programs within schools. Virtually all students compared in this study were from the Freshman Engineering Program (in which all beginning engineering students must first complete the same first-year courses to apply for admission to the specific second-year engineering majors) or the School of Science (which includes prospective mathematics majors), and these two sets of students are comparable in quality and high school preparation. The equations provided the weights needed to combine a given student's scores on the four variables to arrive at a predicted grade point average.

Nonlinear effects of predicted grade point average (PGPA) were investigated by considering quadratic terms. It seemed reasonable to suppose a student would perform better in either of the two programs, C$^4$L or TRAD, the higher this student's PGPA score. However, in spite of the anticipated positive association between academic performance in either of the two programs, C$^4$L or TRAD, and PGPA, it was felt the association might not be linear. For instance, the highest PGPA students might perform better than the moderate PGPA students in either of the two studied programs, but by an amount less than the amount that the moderate PGPA students performed better than the modest PGPA students in either of the two studied programs. A student's major was grouped into the four categories of either engineering, math, science or other. To control for a student's major allowed a distinction to be made between a student's academic performance in calculus due to taking the C$^4$L or TRAD program from a student's academic performance in calculus due to how close this student's major was related to calculus. For instance, it might be the case that a student in mathematics who, because s/he had been accustomed to TRAD type mathematics courses, might actually perform worse than a student in engineering. Or, a freshman engineering student at Purdue may have been exposed to more mathematics courses and have done better in high school courses, traditional or otherwise, and would thus be more likely to perform better in the nontraditional $C^4L$ program.

Gender was also accounted for in the study. To control for gender allowed a distinction to be made between a student's academic performance in calculus due to taking the $C^4L$ or TRAD course from a student's academic performance in calculus due to his/her gender. Gender was controlled for because there was anecdotal evidence provided by the Director of Women in Engineering Program at Purdue that females tended to perform better than males in the group oriented $C^4L$ program as opposed to the more individually (competitive) oriented TRAD program.

In addition to considering just the main effects of predicted grade point average (PGPA), major and gender on a student's academic performance in taking either the $C^4L$ or TRAD program, the possible interaction effects of these three main effects were also considered. It was thought possible that a student's academic performance in calculus might be due to not just to a simple addition of the independent factors of PGPA and major, for instance, but to some dependent combination of these two factors, such as, say, a bright (high PGPA) engineering student performing as well as, but not better than, an average (moderate PGPA) mathematics student.

This study had to contend with missing data. In some cases, students started but did not finish one or the other of the $C^4L$ or TRAD programs, for example. This study included only those students who started and completed the programs. Important questions of why and how many students started, but did not finish, the two types of programs is left for a possible future study.

Although emphasis was placed on the influence of the three factors of PGPA, major and gender on academic performance, the study also looked at the influence of these three factors on which program, $C^4L$ or TRAD, better inspired students to pursue further study in calculus or, more generally, mathematics.

**Statistical Model.** The main statistical model used in the comparison of the $C^4L$ and TRAD programs is an additive multivariate multiple regression model. The basic idea behind the use of this model is that the explanatory variables, such as predicted grade point average or major of a student, are treated as extraneous factors to be controlled for when assessing the type of program, $C^4L$ or TRAD, effect. The primary statistical model used in this study was of the form,

$$Y_{ij} = \alpha_i + \beta X_{ij} + e_{ij}$$

where,

- a type of program effect, $\alpha_i$, is determined for both the $C^4L$ program, $i = 1$, and the TRAD program, $i = 2$, where this effect has accounted for, or has been adjusted for, the various explanatory (extraneous) variables.
- a response $Y_{ij}$, such as the final grade or number of calculus courses, is observed for each student $j$.
- an explanatory (extraneous) variable $X_{ij}$, such as predicted grade point average or major of a student $j$, is used in conjunction with an appropriately

weighted regression coefficient, $\beta$, in an attempt to "explain" the various student responses.

- a residual, $e_{ij}$, measures the difference between the observed and modeled responses.

The statistical model described above is used to determine the estimated difference in the type of program effect, $\hat{\alpha}_{C^4L} - \hat{\alpha}_{TRAD}$, such as, for instance, the difference in average final grades for the $C^4L$ and TRAD students. Two kinds of estimated differences in the type of program effects are calculated. One kind of estimated difference in type of program is *adjusted* for the explanatory (extraneous) variables. The other kind of difference in type of program is *not* adjusted (or is unadjusted) by the explanatory (extraneous) variables (where $\beta$ and $X_{ij}$ are left out of the model here). Comparing the adjusted and unadjusted differences gives some idea of the amount of influence the extraneous variables have on the type of program. For example, if the adjusted difference in average final grades for the $C^4L$ and TRAD students was lower, say, than its unadjusted equivalent, then this would indicate the explanatory (extraneous) variables, such as predicted grade point average, is confounded with the type of program and that the adjustment was necessary.

Both 95% confidence intervals and tests of the estimated difference in the type of program effect, $\hat{\alpha}_{C^4L} - \hat{\alpha}_{TRAD}$, adjusted and unadjusted, were determined. For example, the adjusted 95% confidence interval for the difference in average final grades for the $C^4L$ and TRAD students was,

$$0.40 \pm 0.06.$$

where the grading scale for this study was given as: A = 4.00, B = 3.00, C = 2.00, D = 1.00 and F = 0.00. This means, with 95% confidence, the $C^4L$ students attained, on average, 0.4 of a final grade more than the TRAD students, give or take a standard deviation of 0.6 of a grade. Similarly, a test revealed the chance, or *p*-value, of observing an adjusted difference in average final grades for the $C^4L$ and TRAD students of 0.40, assuming the actual adjusted difference in average final grades is zero, is so small,

$$p\text{-value} = 0.0001,$$

in fact, that it would almost surely be concluded the average difference could not be zero. The complete set of 95% confidence intervals and tests are given in the next section.

The statistical model was, for the most part, analysed using the computer SAS procedure `GLM` ([**27**], [**28**]). However, other parts of the SAS computer package were used, such as the logistic regression procedures `CATMOD` and `LOGISTIC` ([**1**], [**28**]).

## 3. Results and Discussion

Both 95% confidence intervals and tests of the estimated difference in the type of program effect, $\hat{\alpha}_{C^4L} - \hat{\alpha}_{TRAD}$, adjusted and unadjusted, were determined and summarized in Table 1. An interpretation of this summary follows.

**Table 1 Evaluation of $C^4L$ Versus TRAD Programs: Unadjusted and Adjusted Results. Summary of Results.**

| $C^4L$ Versus TRAD Unadjusted Results | 95% CI | Test: $p$-value |
|---|---|---|
| Average final grade | $0.40 \pm 0.06$ | 0.0001 |
| Average final grade for all mathematics courses | no statistical significant differences | |
| Last available overall grade point average | $0.13 \pm 0.05$ | 0.0074 |
| Average number of calculus courses | $0.16 \pm 0.05$ | 0.0038 |
| Average number of noncalculus mathematics courses* | $0.27 \pm 0.07$ | 0.0001 |
| **$C^4L$ Versus TRAD Adjusted Results** | **95% CI** | **Test: $p$-value** |
| Average final grade | $0.42 \pm 0.06$ | 0.0001 |
| Average final grade for all mathematics courses | no statistical significant differences | |
| Last available overall grade point average | $0.09 \pm 0.05$ | 0.05 |
| Average number of calculus courses | $0.24 \pm 0.05$ | 0.0001 |
| Average number of noncalculus mathematics courses* | $0.16 \pm 0.07$ | 0.025 |

The last result in both tables is starred (*) because these results are somewhat misleading, as is explained in greater detail below.

**Average final grade.** Both the adjusted and unadjusted 95% confidence intervals and $p$-value for zero difference in average final grades demonstrate $C^4L$ students earn higher grades in calculus courses; in fact, almost half a grade higher, on average, than the TRAD students.

**Average final grade for all mathematics courses.** There is no evidence, using either the adjusted or unadjusted results, to conclude that there is a difference between the average grades of $C^4L$ and TRAD students in mathematical courses beyond calculus. This suggests that $C^4L$ students are as adequately prepared for mathematics courses beyond the calculus programs, as the TRAD students. We have had much anecdotal evidence from $C^4L$ students who have communicated their displeasure with the traditional lecture, or lecture and recitation, teaching format of math (and other science) classes after taking $C^4L$ calculus courses. This may explain why the two groups of students, $C^4L$ and TRAD, performed at the same academic level in mathematical courses taken after the calculus programs.

**Last available overall grade point average.** The unadjusted, and particularly the adjusted, 95% confidence intervals and $p$-value for zero difference in average overall grade point average demonstrated the $C^4L$ students performed only *slightly* better, on average, than the TRAD students. Again, this suggests that $C^4L$ students are as adequately, possibly slightly better, prepared as the TRAD students for any academic courses beyond the calculus programs.

**Number of calculus courses taken.** Both the adjusted and unadjusted 95% confidence intervals and $p$-value for zero difference in average number of calculus courses demonstrate $C^4L$ students take more calculus courses, on average, than the TRAD students. This suggests that the $C^4L$ program better inspires students, rather than the TRAD program, to pursue further study in calculus. Indeed, the unadjusted 95% confidence interval suggests, that on average, out of 100 courses, the $C^4L$ students take 24 more calculus courses, give or take a standard deviation of 5 courses, than their TRAD counterparts.

**Number of noncalculus math courses taken beyond calculus.** At first, based on both the adjusted and unadjusted 95% confidence intervals and $p$-value for zero difference in average number of noncalculus mathematics courses above, it appeared as though the $C^4L$ students took more noncalculus mathematics courses, on average, than the TRAD students. This suggested that the $C^4L$ program better inspires students, rather than the TRAD program, to not only pursue further study in calculus, but to pursue further study in mathematics courses, in general.

However, these results seemed to be at odds with the raw data, given in Table 2 which seem to show that the average number of noncalculus mathematics taken by the $C^4L$ and TRAD students to be about the same.

**Table 2 Number of mathematics courses taken beyond calculus**

| TRAD Students | Overall | Male | Female |
|---|---|---|---|
| Zero courses | $1763/4431 = 39.79\%$ | $1196/3265 = 36.63\%$ | $567/1166 = 48.63\%$ |
| One course | $2080/4431 = 46.94\%$ | $1607/3265 = 49.22\%$ | $473/1166 = 40.57\%$ |
| Two courses | $431/4431 = 9.73\%$ | $358/3265 = 10.96\%$ | $73/1166 = 6.26\%$ |
| Three or more courses | $157/4431 = 3.54\%$ | $104/3265 = 3.19\%$ | $53/1166 = 4.55\%$ |
| $C^4L$ Students | Overall | Male | Female |
| Zero courses | $81/205 = 39.51\%$ | $47/127 = 37.01\%$ | $34/78 = 43.59\%$ |
| One course | $81/205 = 39.51\%$ | $57/127 = 44.88\%$ | $24/78 = 20.41\%$ |
| Two courses | $18/205 = 8.78\%$ | $11/127 = 8.66\%$ | $7/78 = 8.97\%$ |
| Three or more courses | $25/205 = 12.20\%$ | $12/127 = 9.45\%$ | $13/78 = 16.67\%$ |

As a consequence of this discrepancy, we further examined this aspect of the data in two different ways.

In one analysis, we *excluded* the 1844 students who did not take any mathematics courses beyond calculus and ran the statistical model, described above, for the remaining students. The resulting unadjusted $p$-value of 0.0001 contrasted with the adjusted $p$-value of 0.38 for zero difference in average number of noncalculus mathematics courses taken after the calculus programs. This mixed result tends to support the idea that the $C^4L$ students took about the same number of noncalculus mathematics courses after the calculus programs, on average, as the TRAD students.

In the other analysis, we again excluded those students who did not take any mathematics courses beyond calculus, but, this time, ran a different statistical model, a *logistic* model, for the remaining students. An unadjusted $p$-value of 0.078 results for a test of zero difference in average number of noncalculus mathematics courses. Again, this result tends to support the idea that the $C^4L$ students took about the same number of noncalculus mathematics courses after the calculus programs, on average, as the TRAD students.

## 4. Summary and Conclusions

The results of this study favor the $C^4L$ program over the TRAD program. In answer to the question of which program, $C^4L$ or TRAD, provides a student with a better understanding of the required calculus concepts, we found that:

- The $C^4L$ students earn higher grades in calculus courses; in fact, almost half a grade higher, on average, than the TRAD students.

- The C$^4$L students are as adequately prepared for mathematics courses beyond calculus as the TRAD students.
- The C$^4$L students are as adequately, possibly slightly better, prepared as the TRAD students for any academic courses beyond calculus.

In answer to the question of which program, C$^4$L or TRAD, better inspires students to pursue further study in calculus or, more generally, mathematics, we found,

- The C$^4$L program better inspires students, more than the TRAD program, to pursue further study in calculus.
- The C$^4$L students took about the same number of noncalculus mathematics courses after the calculus, on average, as the TRAD students.

## References

[1] A. Agresti. *Categorical Data Analysis*. Wiley, New York, NY, 1990.

[2] M. Asiala, A. Brown, D. DeVries, E. Dubinsky, D. Mathews, and K. Thomas. A framework for research and curriculum development in undergraduate mathematics education. In A. Schoenfeld, E. Dubinsky, and J. Kaput, editors, *Research in Collegiate Mathematics Education*, pages 1–32, Providence, Rhode Island, 1996. American Mathematical Society.

[3] M. Asiala, J. Cottrill, E. Dubinsky, and K. Schwingendorf. The students' understanding of the derivative as slope. *Journal of Mathematical Behavior*, 16(4):399–431, 1997.

[4] B. Baker and L. Cooley. A schema triad—a calculus example. submitted to *Journal for Research in Mathematics Education*.

[5] J. Bookman. The effects of calculus reform on student performance in subsequent courses. (in preparation) In S. L Ganter (Ed.), *Ten years of calculus reform. A report on evaluation efforts and national impact.* New York, NY:Kluwer Academic/Plenum Publishers.

[6] J. M. Clark, F. Cordero, J. Cottrill, B. Czarnocha, D. J. DeVries, D. St. John, and G. Tolias. Constructing a schema: The case of the chain rule. *Journal of Mathematical Behavior*, 16(4):345–364, 1996.

[7] J. Cottrill. *Students' Understanding of the Concept of Chain Rule in First Year Calculus and the Relation To Their Understanding of Composition of Functions*. PhD thesis, Purdue University, 1999.

[8] J. Cottrill, E. Dubinsky, D. Nichols, K. Schwingendorf, and K. Thomas. Understanding the limit concept: Beginning with a coordinated process schema. *Journal of Mathematical Behavior*, 15(2):167–192, 1996.

[9] B. Czarnocha, E. Dubinsky, S. Loch, V. Prabhu, and Vidaković. D. Students' intuition of area and the definite integral: Chopping up or sweeping out. in preparation.

[10] J. A. Donaldson. A report on calculus i (mathematics 015-156-01) 1993 spring semester. Technical Report Mathematics 015-156-01, Howard University, 1993.

[11] E. Dubinsky. Programming in calculus. unpublished manuscript.

[12] E. Dubinsky. A learning theory approach to calculus. In Z. Karian, editor, *Symbolic Computation In Undergraduate Mathematics Education*, number 24 in MAA Notes, pages 43–55. Mathematical Association of America, Washington, DC., 1992.

[13] E. Dubinsky. *ISETL*: A programming language for learning mathematics. *Communications in Pure and Applied Mathematics*, 48:1027–1051, 1995.

[14] E. Dubinsky. Assessment in one learning theory based approach to teaching: A discussion. In B. Gold, S. Keith, and W. Marion, editors, *Assessment Practices In Undergraduate Mathematics*, number 49 in MAA Notes, pages 229–232. Mathematical Association of America, Washington, DC., 1999.

[15] E. Dubinsky and K. Schwingendorf. Calculus, concepts and computers: Some laboratory projects. In C. Leinbach, editor, *The Laboratory Approach to Teaching Calculus*, number 20 in MAA Notes, pages 197–212. Mathematical Association of America, Washington, DC., 1990.

[16] E. Dubinsky and K. Schwingendorf. Constructing calculus concepts: Cooperation in a computer laboratory. In C. Leinbach, editor, *The Laboratory Approach to Teaching*

*Calculus*, number 20 in MAA Notes, pages 47–70. Mathematical Association of America, Washington, DC., 1990.

[17] E. Dubinsky and K. Schwingendorf. *Calculus, Concepts, and Computers: Multivariable and Vector Calculus*. McGraw-Hill, Raleigh, NC, 1996. Preliminary version.

[18] E. Dubinsky, K. Schwingendorf, and D. M. Mathews. *Calculus, Concepts, and Computers*. McGraw-Hill, Raleigh, NC, 1995.

[19] S. L. Ganter. Calculus renewal: Issues for the next decade. (in preparation) New York, NY: Kluwer Academic/Plenum Publishers.

[20] S. L. Ganter. The impact of ten years of calculus reform on student learning and attitudes. *Association for Women in Science Magazine*, 26:10–15, 1997.

[21] J. F. Hurley, U. Koehn, and S. L. Ganter. Effects of calculus reform: Local and national. *American Mathematical Monthly*, 106(9):800–811, 199?

[22] D. M. Mathews. Time to study: The $c^4l$ experience. *UME Trends*, 7(4), 1995.

[23] D. M. Mathews, M. McDonald, and K. Strobel. Understanding sequences: A tale of two objects. (this volume).

[24] D. M. Mathews, J. Narayan, and K. E. Schwingendorf. Using *maplev* as a mathematical programming language in calculus. In *Proceedings of the Seventh Annual International Conference on Technology in Collegiate Mathematics*, pages 309–313, Reading, MA, 1996. Addison-Wesley.

[25] M. E. Miller. Understanding sequences: A tale of two objects, May 1981. Miller, M. E. (May 1, 1981). Letter to Mr. John Krivacs, Director of Admissions, Indiana University-Purdue University at Indianapolis, from M. E. Miller, Assistant to the Registrar of Purdue University (Betty M. Suddarth), describing the method used for predicting the probability of success of beginning freshman using the PGPA at Purdue University, West Lafayette, IN. (Available from K. E. Schwingendorf).

[26] S. Monteferrante. Implementation of calculus, concepts and computers at dowling college. *Collegiate Microcomputer*, 11(2):95–99, 1993.

[27] J. Neter, M. H. Kutner, and C. J. Nachtsheim. *Applied Linear Models*. Irwin, Homewood, IL, 4 edition, 1996.

[28] SAS Institute Inc., Cary, NC. *SAS/STAT User's Guide, Version 6, Volumes 1 and 2*, 4 edition, 1989.

[29] A. Schoenfeld. In A. Schoenfeld, editor, *Student Assessment In Calculus. A Report of the NSF Working Group On Assessment In Calculus*, number 43 in MAA Notes, pages 31–45. Mathematical Association of America, Washington, DC., 1997.

[30] K. E. Schwingendorf. Assessing the effectiveness of innovative educational reform efforts. In G. Harel and E. Dubinsky, editors, *Assessment Practices In Undergraduate Mathematics*, number 49 in MAA Notes, pages 249–252. Mathematical Association of America, Washington, DC., 1999.

[31] K. E. Schwingendorf and E. Dubinsky. Calculus, concepts and computers: Innovations for learning calculus. In A. Schoenfeld, editor, *Priming the Calculus Pump: Innovations and Resources*, number 17 in MAA Notes, pages 175–198. Mathematical Association of America, Washington, DC., 1990.

[32] K. E. Schwingendorf, J. Hawks-Hoover, and J. Beineke. Horizontal and vertical growth of the students' conception of function. In G. Harel and E. Dubinsky, editors, *The Concept of Function: Aspects of Epistemology and Pedagogy*, number 25 in MAA Notes, pages 133–149. Mathematical Association of America, Washington, DC., 1992.

[33] K. E. Schwingendorf, D. M. Mathews, and E. Dubinsky. Calculus, concepts, computers and cooperative learning, $c^4l$: The purdue calculus reform project. In *Proceedings of the Seventh Annual International Conference on Technology in Collegiate Mathematics*, pages 402–406, Reading, MA, 1996. Addison-Wesley.

[34] K. E. Schwingendorf and G. J. Wimbish. Attitudinal changes of calculus students using computer enhanced cooperative learning, 1994. Paper presented at the Annual Joint Winter Meetings of the AMS and MAA, Cincinnati, OH.

[35] J. Silverberg. Does calculus reform work? In B. Gold, S. Keith, and W. Marion, editors, *Assessment Practices In Undergraduate Mathematics*, number 49 in MAA Notes, pages 245–248. Mathematical Association of America, Washington, DC., 1999.

[36] B. M. Suddarth and S. E. Wirt. Predicting course placement using percentage information. *College and University*, Winter:186–194, 1974.

Mathematics and Physics Section, Purdue University North Central, Westville, Michigan 46391-9528

*Current address*: Mathematics and Physics, Purdue University North Central, Westville, Michigan 46391-9528

*E-mail address*: kschwing@purduenc.edu

Department of Statistics, School of Science, Purdue University, West Lafayette, Indiana

*E-mail address*: mccabe@stat.purdue.edu

Mathematics and Physics Section, Purdue University North Central, Westville, Michigan 46391-9528

*Current address*: Mathematics and Physics, Purdue University North Central, Westville, Michigan 46391-9528

*E-mail address*: jkuhn@purduenc.edu