

33. Methodology, Findings, and Conclusions

Conard White

A total of 59 teacher teams were organized from participating school districts beginning in the 1996/1997 school year through the 2003/2004 school year. The number of teachers in each team ranged from two to as many as four, and in each case represented more than one subject matter area. Teachers volunteered to participate in the program, to work together across curricular areas to develop interdisciplinary instructional modules, and to pilot test the modules within their regularly scheduled classes. Data was collected on nearly 2800 students across 67 classes. In most cases, the data from a team represented a single "class," as the same instrument was administered to all participating students. The data from seven teams, however, were separated by subject matter because different test instruments were used.

Teachers received special instruction in team dynamics, integrated curriculum models, authentic assessment strategies, teaching models, and use of state standards in developing learning modules. Instructional modules developed by each team met specific criteria:

- Have a duration of three to six weeks
- Be based in specific standards taken from the Illinois Learning Standards
- Use at least two of the integrated curriculum models presented
- Encompass content across MSTE disciplines
- Include two or more authentic assessment strategies as well as pretests and posttests

Each module focused on concepts common across disciplines. These included common curricular content, common cognitive content, and common skills. Modules encompassed such concepts as problem solving, oral, written, and graphic communication, computer skills, critical thinking, computation, and measuring.

During the first project year, the research model employed a quasi-experimental, nonequivalent control group design. Treatment groups were constituted in one of two ways. In two high schools, the treatment group consisted of students common to each of the disciplinary areas of the team. During the school day, each student in this group came in contact with each teacher in the team. In two high schools, the treatment group was not composed of common students; however, the teachers had common planning periods and were able to employ team-teaching strategies in each other's classes. Control groups were selected from classes in the same disciplines as the treatment groups but had no teachers in common nor did they employ team-teaching strategies. Posttest data were collected from both groups at the conclusion of each module. An analysis of covariance using a generalized linear model was used to test for differences in mean posttest scores between the treatment and control groups. Pretest scores were employed as a covariate to control for the effects of student ability and prior knowledge. The research hypothesis tested using this model was:

H₁: There is no significant difference in adjusted mean achievement scores between the treatment and control groups.

As a result of research done in the previous year, it was found that the use of control groups to adequately assess achievement gain attributed to the instructional module was not practical in the school setting. It was very difficult to locate groups of students that were comparable to those participating in the study, to gain cooperation of additional teachers not associated with the program, and to coordinate the testing of possible control groups within the timeframes of each of the modules. It was decided to eliminate the use of student control groups, although this reduced the effectiveness of the research model.

In order to test for student gain in achievement, each team, as a part of its design of instructional modules, developed a traditional test instrument. This instrument had to be designed to measure the same instructional objectives taken from the Illinois Learning Standards as those incorporated into the instructional module. These tests were designed by the team instructors, approved by the project research associate, and administered at the beginning of module instruction to assess students' prior knowledge of the subject matter. At the end of the module, the test was again used to measure student accomplishment.

A paired-difference t-test was employed to test for significant differences between the pretest and posttest mean scores. A significance level of 0.05 was selected to distinguish a significance. The hypothesis being tested here is:

H₂: There is no significant difference between mean pretest and mean posttest scores.

To analyze the effects of gender and race/ethnicity on posttest scores, a generalized linear covariance model was employed. Pretest scores were used as a covariate to control for the effects of prior knowledge on the part of the students. A significance level of 0.05 was used in these tests also. The hypotheses being tested here are:

H₃: There is no significant difference in adjusted mean posttest scores between male and female students.

H₄: There is no significant difference in adjusted mean posttest scores among students in differing racial/ethnic categories.

Limitations

Team members designed the instruments used to assess the effectiveness of the instructional modules. They were given instruction in developing test items to measure specific objectives. The completed instrument met these requirements:

- Measure instructional objectives contained within the module.
- Include items that measure content in each subject matter area.
- Include a variety of item types.

The instruments developed by the teams were monitored by project staff, and a variance in quality was noted. No attempt at assessing the validity was made, other than ensuring that the test items were directed toward state standards. Reliability checks on the test instruments were not required. Each instrument is assumed to have been reliable, since the same test was used before and after the instructional module.

The initial study, conducted by the project staff in academic year 1997/1998, used a more sophisticated research model employing equivalent control groups. As a result of consultation with participating teachers and school administrators, it was determined that the use of control groups was not practical in a public school setting. It was difficult to locate groups of students that were equivalent to those participating in the study. If equivalent sections of classes were taught by project teachers, it was thought that “cross-over” might occur between the control and experimental groups. It was difficult to find teachers not associated with the project who would cooperate by allowing testing of their students. Project staff determined there was no alternative to eliminating use of control groups, although it greatly reduced the power of the research model.

Performing research in schools is a complex endeavor, ranging from the level of administrative support to the logistics of collecting data from teachers in the prescribed format. There are limitations in large and complex districts. For example, this initiative began using control groups the first two years, but the following years we were unable to schedule control groups. The burden was too great for the district to coordinate. It took considerable effort to get teachers to report the pretest and posttest scores on the forms provided, which entailed much follow-up. In spite of the research limitations and the “messy” business of social research, the research model and results reported here have positive implications.

Findings

The project’s website contains all data and statistical procedures employed to evaluate student achievement, from 1997 through 2003: www.strategicalliance.niu.edu.

Summary and Conclusions

During the seven years of the project, 59 teacher teams were trained; they developed more than 102 instructional modules and piloted them in the classroom. These teams consisted of teachers from five Rockford high schools, six Rockford middle schools, and four high schools in northern Illinois. Data were collected from 2794 students.

In order to test for student gain in achievement, each team, as a part of the design of instructional modules, developed a traditional test instrument. This instrument was designed to measure the same instructional objectives taken from the Illinois learning standards as those incorporated into the instructional module. These tests were designed by the team instructors and administered at the beginning of module instruction to assess students’ prior knowledge of subject matter. At the end of the module, the test was used to measure student accomplishment.

Figure 1 displays the results of the posttest administered at the close of the instructional module. The chart indicates the frequency of mean class scores in categories roughly associated with ABCDF grade ranges. That is, 3 of the participating classes had a class average in the “A” range, and another 16 classes had an average in the “B” range. The fact that 28% of the classes

participating (16 out of 67¹) had averages of “A” or “B” is outstanding. The largest number of classes (23 or 34%) had an average in the “C” range. Sixty-two percent of the classes represented in the study had “final exam” class averages in the “C” or above range. The data clearly indicate that high levels of learning were taking place within the participating classrooms.

Gain in achievement of each of the participating classes can be shown by comparing the posttest with the corresponding pretest class averages. A significant gain in achievement was demonstrated in 67 of the 68 classes represented in the study. One class showed a slight but insignificant gain. Those data are displayed in Figure 2.

As can be seen from the data, six classes experienced a gain in achievement of over 200% during the course of the instruction. In addition, 25% of the classes at least doubled their achievement, while over half experienced a gain that exceeded 50%.

In order to summarize the findings of the project, one must examine the 1997/1998 year separately from the others. A fundamental change in research design took place after that year. The major thrust of year one was to determine if the use of teacher teams in conjunction with integrated curriculum could demonstrate an increase in student achievement when compared with a traditional separate curriculum approach. Treatment groups, composed of students enrolled in classes incorporating an integrated curriculum, etc., were compared with control groups of students enrolled in regular separate subject areas. The success of this year would form the basis for further study in subsequent years.

¹ There were 68 classes. A percent correct could not be determined for one class that did not report the number of points possible.

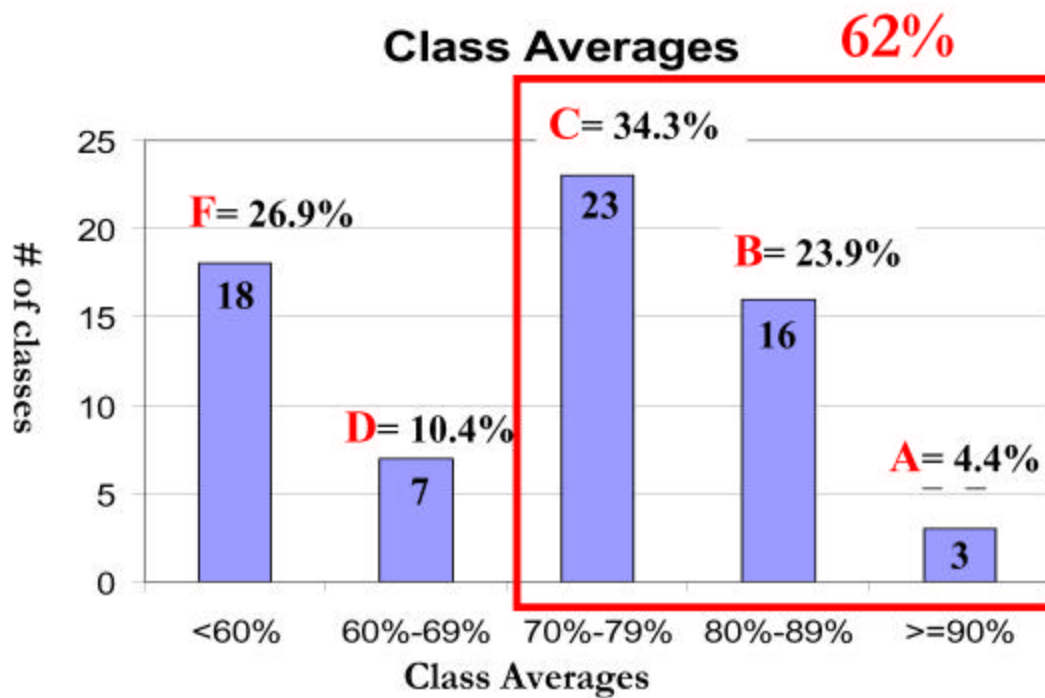


Figure 33.1 Class Averages

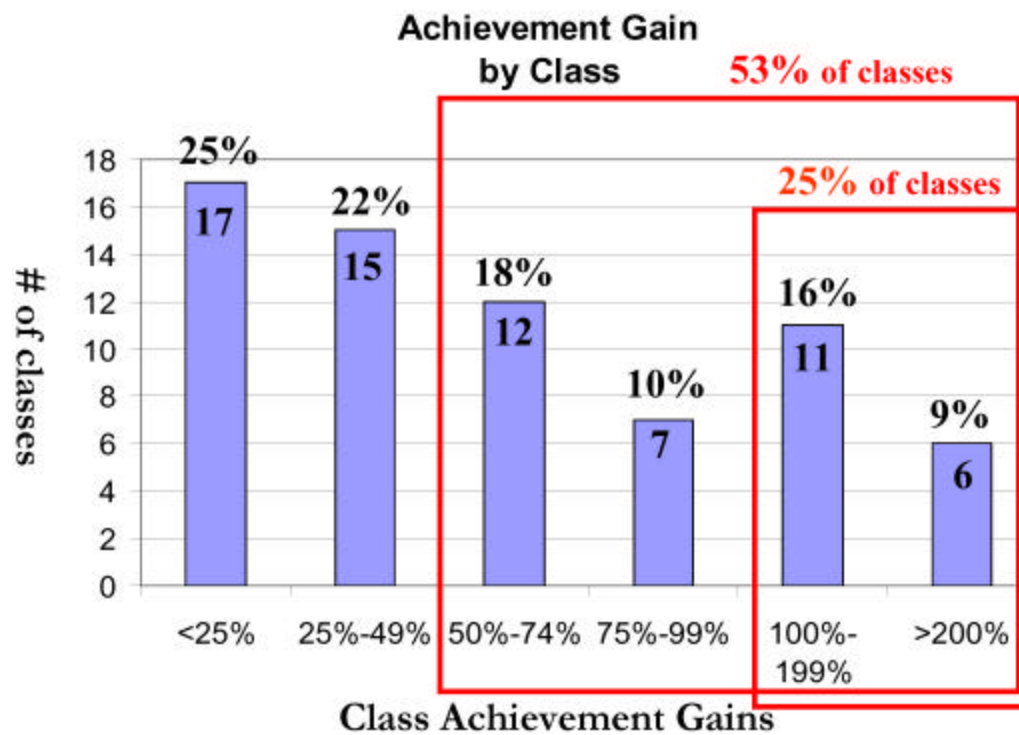


Figure 33.2 Achievement Gain by Class

Data collected during the first project year indicated that the integrated curricular/teacher-team method was viable. Data from four of the nine first-year teams exhibited a significantly higher gain in achievement in the classes using the integrated curricular approach when compared to the classes using the traditional separate approach. The data from the remaining five teams showed no significant differences between the treatment and control groups. None of the teams experienced outcomes where the integrated curricular approach was less effective than the separate approach. The conclusion is that the use of teacher teams in conjunction with integrated curriculum, etc. can improve student achievement over a traditional, separate-curriculum approach.

Subsequent project years focused not only on student achievement gain over the course of instruction but also on the effectiveness of the integrated curricular/teacher-team approach with regard to gender and race/ethnicity. In almost every case, data collected during those years indicated that a significant gain in achievement took place over the course of instruction.

Of the 59 teams reporting during these years, only seven showed a significant difference in achievement scores between male and female students. Of those seven teams, five showed that females scored higher than males, while the reverse was true in the remaining two teams. No interviews were conducted with teams to determine reasons for these differences. However, it is interesting to note that the only module subject areas where males outscored females were architecture and physical education. All of the modules where females topped males in achievement featured the subject matter area of English.

Ten teams reported data that displayed a significant difference in achievement among the categories of race/ethnicity. In six of those teams, Caucasian students displayed higher achievement scores than students in other categories. In three teams, Asian students outscored students in the other categories. No additional information was available that might explain the differences.

Recommendation

The models, products, and results reported here have been evaluated and provide evidence of improved student achievement in mathematics, science, technology, and English at the secondary level. Integrated MSTE curricula, new teaching models and strategies, improved traditional tests, and new performance assessments may improve student success. However, the studies should be replicated with tighter controls to enrich and strengthen the empirical evidence. See the website: www.strategicalliance.niu.edu, for supplementary information.