THE EVALUATION OF FEDERAL PROGRAMS IN EDUCATION

Garrett K. Mandeville, University of South Carolina

Introduction

Beginning in about 1973 the federal government developed an interest in the evaluation of the effects of monies provided to support public education. In fact, loosely written evaluation requirements were included in much of the legislation such as The Education Amendments of 1974 to Title I of the Elementary and Secondary Education Act. One of these federal programs, initially referred to as Title III and later amended to be known as Title IV-C, focuses on the development of innovative programs.

Title IV-C projects generally run for three and one-half years, with the first six months being used to complete the proposal including the evaluation design, the latter being prepared by an external evaluator hired by the project. Once the proposal is approved, project staff generally spend a year developing the products, and a second year field testing them. Although the evaluator has some limited involvement in monitoring these developmental activities, his or her primary task is to conduct the evaluation during the third year, when the project undergoes a full implementation.

Although educational evaluators are generally competent individuals with appropriate backgrounds in research and evaluation, measurement and statistics, at times projects evolve so that it is difficult or impossible to conduct the evaluation as originally planned. It has been at times like this that the writer has been called upon as a consultant to either (1) write the code to conduct the analysis as originally planned, or (2) identify a reasonable alternative analytic strategy and code it. This paper provides a description of a project for which the writer served as a consultant, and the algorithms developed to perform the analyses.

A General Description of CBE Programs

The program to be described reflects a recent movement in education towards assuring some (possibly minimal) level of competency; these programs are often labeled as Competency Based Education (CBE) programs. In this approach, local educators develop lists of competencies which they feel should be mastered by all students. Next they develop instructional materials, often based on a mastery learning model, to teach students these competencies, and tests to measure student attainment of the competencies. Each test usually deals with a single competency and the tests are often scored as "pass" or "fail" using subjective or empirically derived criteria. Most of the tests are objective, the items being presented in a multiple choice format. There are usually multiple forms of these tests. One form is used as a pretest to identify the initial level of performance of each student. This allows for the possibility that, if a student has already achieved the competency, he or she may be assigned to more advanced or enrichment instructional activities. Other forms of these tests are used after instruction to determine whether the student requires remedial work. These tests are often referred to as criterion referenced tests (CRTs).

Since these programs are usually implemented on a school-wide basis, control groups formed by randomization are impossible. Rather, comparable schools are identified and some form of a nonequivalent control group design is prescribed in the proposed evaluation scheme.

The Specifics of the Middle School CBE Program

A middle school CBE project grant was awarded to a suburban school district in a major southeastern metropolitan area during the summer of 1977. Subject areas in which the CBE instruction was to take place were language arts, mathematics, science and social studies. Competencies in each of these four subject areas for the first and second semester of the 6th, 7th and 8th grades were identified. The following table indicates the number of competencies for each grade, semester and subject area.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Semester</th>
<th>Subject Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>1</td>
<td>Lang, Math, Science, Soc. Stud.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4, 10, 3, 3, 3</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3, 4, 5, 3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>7, 7, 4, 3</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>3, 4, 5, 9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4, 5, 5, 9</td>
</tr>
</tbody>
</table>

Note: Five competencies are not included in the above counts because the associated tests were not objective and they are not dealt with in this paper.

CRTs were then written for each competency. The number of test items per competency ranged from two to 20 with the exception of three competencies which had to be hand scored and were simply coded pass-fail. Mastery criteria were set for each CRT, with a student required to answer approximately 70-80% of the items correctly to demonstrate mastery.

Original and Post Hoc Middle School Evaluation Design

The original evaluation design was to utilize a control school which was similar to the project school on selected demographic characteristics and previous achievement. All students in both schools were pretested with one form of each of the four subject area CRTs. Presumably, assignment to CBE classes in the project school was to be based on these test scores, thus allowing for identification of appropriate control students, i.e., those students in the control school failing to master the test. The analysis would then be reasonably simple, with first and second semester post CRTs and standardized tests (usually required in these evaluation designs) to serve as the dependent variables. Of course, the experimental and control groups would change for the different subject areas and the status of a project
The major component of the substitute strategy was to categorize project students into the four cells of a two dimensional table based on (1) whether or not they had demonstrated mastery of the subject at pretest, and (2) whether or not they were assigned to a CBE class in that subject. The next step was to match control students to project students based on pre-instructional (spring 1979) performance on the Comprehensive Test of Basic Skills (CTBS), a standardized test which is annually administered district-wide. For clarity, the following table describing the cells in the fourfold table is presented.

### Table 2
**Description of Cells in Fourfold Table Based on Mastery and Classroom Assignment**

<table>
<thead>
<tr>
<th>Pretest Status</th>
<th>Non master</th>
<th>Master</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell 2</td>
<td>Cell 1</td>
<td>Cell 4</td>
</tr>
<tr>
<td>Cell 3</td>
<td>Regular CBE</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** If the project had been implemented as intended, Cells 2 and 3 would have been empty.

For purposes of the evaluation, the most important group of project students were those who had not mastered the subject and who received CBE instruction in the subject, i.e., those in cell 1. The next most important group of project students were those nonmasters who did not receive the CBE instruction. It was felt by project staff that these students in cell 2 might possibly be positively affected by some general school influence. Students in cell 3 were next in priority with little expectation for project vs. control student differences for those in cell 4. Thus, to whatever extent possible, control student matches for project students were to be identified according to the above priority scheme. It should be noted that the matching had to be done independently for each of the four subject areas. However, since project staff indicated there was little movement in and out of CBE classes between semesters, first semester classroom assignments were used throughout. The correlated t test is the obvious comparison statistic and was conducted separately for each of the four cells. The dependent variables were the first and second semester CRT posttest scores (raw scores rather than "pass-fail" scores were used) in the subject area and the appropriate subtest score from the spring 1980 administration of the standardized test, the CTBS. In addition to the raw scores for each competency, total subject area CRT scores were also analyzed.

In this section, some of the coding utilized to efficiently analyze the data will be presented. The initial consideration was that the data were prepared on punch cards which contained student raw responses to each subject area CRT. Responses to each CRT were on individual records. Since there were roughly 175 student records for each subject area, semester, grade, and school, about 8400 test protocols required scoring. Obviously an efficient scoring algorithm was needed. The solution used was to precede each set of records for a new test with a record containing the test key and prepared in the same format as the response records. These records were identified as test keys by punching them with a dummy code. (The five column ID field was set to 99999; although a blank ID field will work, it is dangerous in that an occasional stray blank card or poorly edited card with a blank ID field will find its way into the deck.) The general algorithm then is simply as follows:

**Solution Via SAS**

```
ARRAY RAW R1-R60; ARRAY KEY K1-K60;
ARRAY SCORED S1-S60; RETAIN K1-K60 FLAG;
INPUT(X1-X60)(60*1.) ID 68-72 TESTID 77-79;
IF ID=99999 THEN DO;
	DO OVER RAW; KEY=RAW;END;
	DELETE; FLAG=1; END;
	DO OVER RAW; SCORED=0;
	IF RAW=KEY AND KEY NE • THEN SCORED=1;
END;
TOT=SUM(S1-S60);
IF FLAG=1 THEN DO;
	PUT 'SCORING CHECK FOR NEW TEST' TESTID=/
	‘RAW ’ K1-K60/
	‘KEY ’ K1-K60/
	‘SCORED ’ S1-S60 TOT= ;
	FLAG=0; END; Solution Via SAS
```

Other variables not presented in the above example were also included in the INPUT statement. Among them was a code indicating whether a student in the project school was enrolled in a CBE class. Another complexity not reflected above was that initially eight different data sets were created, one for each subject area and semester combination. When these 8 data sets were merged (by ID), an 8 bit integer indicating the presence or absence of the 8 sets of test scores was created. Later it is referred to by the variable name SCORES.

Next records containing the CTBS pretest (for matching) and posttest scores were input...
and merged with the above. Finally, the cards were read which contained pretest mastery indicators for each of the four subject areas (COM1-COM4). Now we are prepared to attack the main problem of matching controls to experimental students giving priority to the smaller numbered cells. The variables not previously defined are as follows:

CBL-CB4: project student in/not in (1/0) CHE class in subject areas 1-4. (Note: area one is language, two is mathematics, etc.)

COM1-COM4: project student demonstrated/not demonstrated (1/0) mastery of subject area 1-4 at pretest.

CL1-CL4: cell designator (1,2,3,4) to which student is assigned for given subject area.

(See Table 2)

ENT1-ENT4; pretest CTBS test scores used for the above.

W1-W4: Interval widths for grouping pretest scores to be considered "matched".

STR1-STR4: An index of the stratum in ENT that the assigned student was assigned when W4 is the basis for the frequency distribution.

TRT: treatment code

The algorithm presented below was applied to each grade separately since, when the matching had been accomplished, it was followed with coding to perform the analysis which was grade specific.

Assume that the variables STRL and CLI have been renamed to STR and CELL respectively. Call the previous data set TEMP3.

DATA TEMPl; SET TEMP3; IF TRT=1; *EXPTLS
EL=R11; E2=R12; E3=R13; E4=R14;
K5=SUB(DR R1-R16); B5=CTBPOS1;
YE=ENT1; IDE=ID;
DATA TEMP2; SET TEMP3; IF TRT=1; *CONTROLs;
C1=R11; C2=R12; C3=R13; C4=R14;
C5=SUM(OF R1-R14); C6=CTBPOS1;
TC=ENT1; TDG=TD;

At this point, an algorithm was employed which compared the experimental and control frequency distributions on STR and randomly eliminated from either group until the two frequency distributions "conformed." The two reduced data sets were then merged to conclude the artificial matching process. The following coding is completely general and was included as a MACRO in the operational version of the coding.

PROC MEANS NOPRINT DATA=TEMP2; BY STR; VAR YC;
OUTPUT OUT=CON N=FRC;
PROC MEANS NOPRINT DATA=TEMPl; BY STR; VAR YE;
OUTPUT OUT=EXP N=FRE;
DATA ALLFREQ; MERGE EXP(=YES3) CON (IN=YKSC);
BY STR; IF STR=. THEN DELETE;
YES1=0 THEN FRE=0; IF YESC=0 THEN FRC=0;
DIFF=FRE-FRC;
DATA TEMPl;

MARG TEMP1(IN=YESI) ALLFREQ END=EOF; BY STR;
RETAIN KT KTNEED; DROP KT KTNEED TOTC KEPT FRE FRC DIFF;
IF STR=. THEN DELETE;
IF YES1=0 THEN GO TO LAB3;
IF FIRST STR THEN DO; TOTC+FRE; KT=O;
KTNEED=FRE-DIFF*(DIFF GT O); END;
KT1; IF KT LE KTNEED THEN DO; KEPT+1;
OUTPUT; END;
LAB3: IF EOF THEN PUT 'KEPT' KEPT 'OUT OF TOTAL NON BLANKS' TOTE=;
DATA TEMPl;
MERCZ TEMPl(IN=YES2) ALLFREQ END=EOF; BY STR;
RETAIN KT KTNEED; DROP KT KTNEED TOTC KEPT FRE FRC DIFF;
IF STR=. THEN DELETE;
IF YES2=0 THEN GO TO LAB3;
IF FIRST STR THEN DO; TOTC+FRE; KT=O;
KTNEED=FRC+DIFF*(DIFF LT 0); END;
KT1; IF KT LE KTNEED THEN DO; KEPT+1;
OUTPUT; END;
LAB1: IF EOF THEN PUT 'KEPT' KEPT 'OUT OF TOTAL NON BLANKS' TOTE=;
DATA TEMPl;
MERCZ TEMPl(IN=YES3) ALLFREQ END=EOF; BY STR;
RETAIN KT KTNEED; DROP KT KTNEED TOTC KEPT FRE FRC DIFF;
IF STR=. THEN DELETE;
IF YES3=0 THEN GO TO LAB1;
IF FIRST STR THEN DO; TOTC+FRE; KT=O;
KTNEED=FRE-DIFF*(DIFF = 0); END;
KT1; IF KT LE KTNEED THEN Do; KEPT+1;
OUTPUT; END;
LAB1: IF EOF THEN PUT 'KEPT' KEPT 'OUT OF TOTAL NON BLANKS' TOTE=;

The algorithms noted above are completely general and was included as a MACRO in the operational version of the coding. The final stage was to calculate the dependent t statistics. The following simple
solution, which may include more output than some would like, is presented without comment.

```plaintext
ARRAY D D1-D6; ARRAY E E1-E6; ARRAY C C1-C6;
DO OVER D; D=E-G; END;
CELL=FLOOR(CELL); *SEPARATE ANALYSES EACH CELL;
PROC SORT; BY CELL;
PROC MEANS NOPRINT; BY CELL;
VAR D1-D6 E1-E6 C1-C6;
OUTPUT OUT=GREAT N=NSIZE
MEAN=MD1-MD6 ME1-ME6 MC1-MC6
STD=STD1-STD6 STD1-STD6 STD1-STD6
STDER=SE1-SE6 SE1-SE6 SE1-SE6
T=TDIFF1-TDIFF6 PRT=PROBT1-PROBT6;
PROC PRINT;
```

**Summary**

This is only one example (but one of the more complicated ones) of a situation in which the writer has capitalized on the exceedingly flexible data manipulation capabilities of SAS. Before about 1973, the writer would have attacked a problem such as this by writing a FORTRAN program to perform the above analyses. The investment of time and effort would have been, conservatively estimated, twice to three times that needed to develop the SAS strategy. We are indeed fortunate to have such an advanced and ever improving software package at our fingertips; pity the poor researchers, evaluators, analysts, etc., who are not so fortunate.
ECONOMETRICS AND TIME SERIES

Session Leader:

Robert Parks, Washington University