

Evaluation of an Elementary School Social Studies Program

HOWARD A. SULKIN *
ROBERT W. PRANIS

THE challenge involved in introducing a new program in a traditional elementary school curriculum is to write material that will capture the interest and imagination of children and provide them with a significant educational experience. Once this challenge has been met, there remains the problem of evaluating such a program, through tests that are a valid measure of the learning experience. Ideally, such tests aid in developing similar programs or in revising the program under study.

Since 1960, a project has been under way at the Industrial Relations Center of the University of Chicago to develop and test elementary school economics curricula for the fourth, fifth, and sixth grades.¹ The project's goal has been to write program material appropriate for students of these grades while introducing them to economics as a social science discipline. The results of this project will interest not only those concerned with social science education at the elementary school level but also those concerned with curriculum evaluation in general.

The Curriculum

The fourth-grade program (Elementary Economics I) and the fifth-grade program (Elementary Economics II) have the same structure but different content. Each pro-

¹ The sixth-grade program, ESE III, was not included in this study.

gram takes approximately 20 to 25 class hours, or about six weeks, to complete, and is ordinarily given during the social studies hour. Each program is independent of the other: ESE I is not a prerequisite for ESE II, nor does ESE II necessarily have to follow ESE I, although it is assumed that a student should reach a better understanding of economics by taking both programs in sequence.

The principal objective of both programs is to develop understanding of the economic processes of consumption and production: ESE I concentrates on consumption; ESE II emphasizes production. Within the concept of consumption, ESE I focuses on the topics of wants, work, money, and consumption and savings. Examples of the kinds of understandings ESE I students are expected to develop are:

Wants. The people of a nation want an unlimited number of goods and services. Because a nation's productive resources are limited, it can produce only a limited number of goods and services.

Work. Work is the use of physical and

NOTE: The authors are indebted to the Charles Stewart Mott Foundation for the support of this project and to James Murray and Barbara Ireland of the Industrial Relations Center, for their help in the preparation of this article.

* Howard A. Sulkin, Director, Evaluation Research, and Robert W. Pranis, Research Associate, both of the Industrial Relations Center, University of Chicago, Chicago, Illinois

mental energy for the purpose of producing goods and services. Many people obtain satisfaction through work.

Money. Money is anything that is commonly accepted as payment for goods and services. Money which by law must be accepted as payment for goods, services, and monetary debt is called "legal tender."

Consumption and Savings. Consumption is the process of using goods and services to satisfy wants. People who use goods and services are consumers. Everyone is a consumer.

In ESE II, the fifth-grade program on production, some of the economic ideas are:

People organize businesses to make a profit.

To produce goods or services efficiently, labor, materials, and equipment must be carefully selected and efficiently used.

Marketing is the process of: (a) transporting goods to places where they will be available to consumers, and (b) selling goods to consumers.

Production and marketing help consumers by: (a) providing people with jobs for which they receive money, and (b) producing goods and services which people can buy.

The material for each program consists of a student text, a student workbook of exercises and problems related to concepts introduced in the text, and a teacher's manual. The manual contains a detailed outline of the program and includes suggestions for discussion topics and supplementary activities.

Evaluation

The data reported here were collected as part of extensive field-testing. The evaluation study attempted to answer two basic questions about the program: (a) Did the students learn the important economic ideas? (b) Did they remember these ideas for a significant length of time, that is, for at least one year?

The two programs were evaluated independently of each other, with different groups of students. The design of the study was a simple non-equivalent control-group

design (Campbell and Stanley, design number 10).² Pre-program and post-program test scores were obtained for an experimental group (students taking the program) and for a control group (students not taking the program, but like the experimental group in other respects).

The experimental and control treatments can be summarized as follows:

	Year 1		Year 2	
	Experimental	Control	Experimental	Control
Pretest	x	x	x	x
Program	x			
Post-test	x	x	x	x

Six elementary schools in a Midwestern city school system participated in the study. During the first year of the study, the sample comprised students of 24 classrooms: 12 fourth-grade classes (ESE I) and 12 fifth-grade classes (ESE II). Six of the classes for each program were experimental, and six were control. For each program, one experimental and one control class were drawn from each of the six schools. This was done to control, at least partially, for possible differences among schools. In this school system the students were from families considered to be lower middle to middle class in socioeconomic status.

The schools did not practice homogeneous classroom grouping with respect to academic ability. All classes of any one grade level had the same formal curriculum except for the experimental classes, which had the ESE programs. The classrooms selected for the evaluation study, as part of either the control group or experimental group, were not chosen by any explicit criteria other than the teacher's willingness to cooperate.

Only students for whom four test scores were available (first pre- and post- and second pre- and post-) were included in the study. The same pre- and post-tests were

² Donald Campbell and Julian C. Stanley. "Experimental and Quasi-Experimental Designs for Research on Teaching." In: N. L. Gage, editor. *Handbook of Research on Teaching*. Chicago: Rand McNally & Company, 1963. Chapter 5. pp. 171-246.

administered to the experimental and control groups exactly one year from their first administration. The control and experimental groups for both programs did not receive any formal instruction in economics after taking the program. The numbers of subjects whose test scores were included in the final sample were:

ESE I experimental group:	N = 58
ESE I control group:	N = 32
ESE II experimental group:	N = 123
ESE II control group:	N = 100

Two types of analysis might have been used in this study: classroom as a unit (that is, group mean) or N-odd student scores of the group. It was decided to use N-odd student scores of the group as the unit of analysis for the following reason. The makeup of the first year classrooms did not stay constant to the next year. Some students left the school districts, while a number of other students when passed to the next grade were recombined into a new class structure, having some students who had taken the program and other students who had not taken the program.

Test Instruments

The pre- and post-tests for each program were parallel forms, constructed to match, item for item. Half of each experimental class and half of each control class in the evaluation of each program received one form as a pretest and the other form as a post-test, while the remaining half of the students received the tests in reverse order. In the second year of the study, each

student received the forms in the same order as he received them in the first year.

The reliability estimates of the parallel forms, a measure of the equivalence of one form to the other, are given in Table 1. Test order was not found to be a significant factor. The reason for having two forms for each program's test is that any test-retest effects (or the possibility that a student's past experience with a test will raise his subsequent scores) are reduced. With the two parallel forms, a student would have taken the same test four times, with an interval as short as six weeks between some of the administrations.

The tests were intended to assess the student's knowledge of the economic concepts taught in the programs. The questions were multiple-choice (three choices), focusing on definitions and vocabulary covered in each program. There were 27 items on the ESE I instruments and 24 on the ESE II instruments. Examples of items are given below:

Examples of Fourth-Grade Test Items (ESE I)

Human wants are:

- Easily satisfied
- Always changing
- Usually the same for most people.

Doctors, barbers, and teachers:

- Do not produce services in their work
- Produce goods
- Produce services.

When one baker makes a cake and another baker puts on the frosting, this is called:

- Division of labor
- Apprenticeship
- Craftsmanship.

Examples of Fifth-Grade Test Items (ESE II)

Consumption is the process of:

- Producing goods and services
- Buying goods
- Using goods and services.

People work to obtain money so they can:

- Satisfy all their wants
- Exchange it for their labor
- Buy goods and services.

		ESE I (Fourth grade)			
		Experimental		Control	
		Year I	Year II	Year I	Year II
Pretest		.58	.60	.51	.62
Post-test		.77	.81	.48	.65
		ESE II (Fifth grade)			
		Experimental		Control	
		Year I	Year II	Year I	Year II
Pretest		.53	.57	.47	.59
Post-test		.72	.79	.64	.69

Table 1. Reliability Estimates (KR-21):
ESE I and ESE II Tests

The process of combining labor, materials, and equipment in order to make something is called:

- Consumption
- Production
- Distribution.

Results and Conclusions

The results are summarized in Figure 1 in terms of experimental- and control-group means for each program. While the statistical treatment of the data did not include analysis of means, this graph of mean scores shows the general pattern of test scores in the most simple and straightforward fashion. The statistical evaluation was done by a two-way analysis of covariance,³ with data for each program analyzed separately (see Table 1).

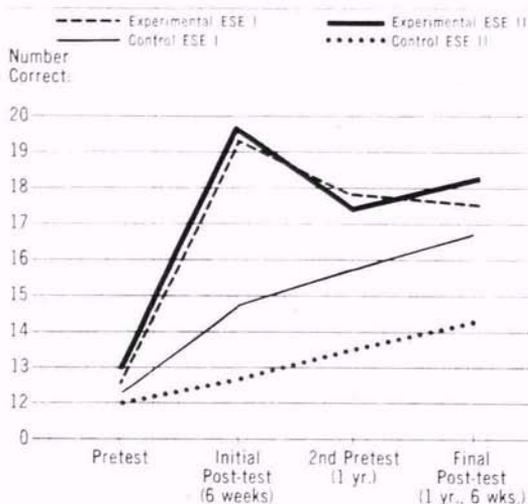


Figure 1. ESE Group Means Before and After Program

The experimental hypotheses were:

- Students who participated in the programs (experimental groups) would have higher test scores than students who did not participate in the programs (control groups) when these students were tested immediately upon the conclusion of the programs six weeks later.

³ In our statistical treatment, the first pretest was defined as the independent variable, with the following three tests covarying on the initial test.

- When the same students were tested exactly a year later, the experimental groups would still have higher test scores than the control groups if the experimental students retained what they had learned.

	Mean Program	Mean Control	S.D. (w/in)	Regression Coefficient between Pre- and Post-	S.E. (Beta)
Pretest Initial	12.61	12.16	3.81	—	—
Post-test	19.38	14.84	4.11	.66	.09
Second Pretest	17.76	15.72	4.17	.58	.10
Final Post-test	17.66	16.78	4.36	.61	.11

Table 2. Means, Regression Coefficients, and Standard Errors: ESE I (N = 90)

Both hypotheses held for ESE II⁴: on the first and second post-test scores (see Tables 5 and 6) the experimental and control students differ significantly ($p < .0001$). For ESE I (see Tables 2 and 3) the first hypothesis was supported ($p < .0001$) while the second, concerning retention, was not. Analysis of experimental and control differences on the second post-test showed that the difference was not at a significant level, although on the test given six weeks earlier, the second pretest, there was a significant difference between experimental and control groups ($p < .025$).

Source	Multivariate F-ratio	df	p (<)
Test Form	1.41	3, 83	.25
Treatment (Eliminating Form)	16.56	3, 83	.00
Form X Treatment	1.50	3, 83	.22

Table 3. Analysis of Covariance on Mean Vectors: ESE I

A glance at the graph of mean scores for ESE I (Figure 1) helps to clarify this non-significant difference on ESE I second post-test scores. It appears that this small difference between experimental and control groups is due not so much to low scores for experimental students (forgetting), but to

⁴ "Multivariate," by Jeremy Finn, running on the IBM-7094, was used to perform the statistical analyses.

Source	Variate	Mean Square	Univariate F-Ratio	df	p (<)	Step-down F-ratio	df	p (<)
Test Form	Initial Post-test	22.61	2.10	1, 85	.15	2.10	1, 85	.15
	Second Pretest	2.24	.18	1, 85	.68	.26	1, 84	.62
	Final Post-test	.51	.04	1, 85	.85	1.88	1, 83	.18
Treatment (Eliminating Form)	Initial Post-test	371.95	34.47	1, 85	.00	34.47	1, 85	.00
	Second Pretest	65.85	5.21	1, 85	.03	1.34	1, 84	.25
	Final Post-test	7.68	.56	1, 85	.46	9.90	1, 83	.00
Form X Treatment	Initial Post-test	21.99	2.04	1, 85	.16	2.04	1, 85	.16
	Second Pretest	1.18	.09	1, 85	.76	1.87	1, 84	.18
	Final Post-test	.04	.00	1, 85	.96	.58	1, 83	.45

Table 4. Multivariate Analysis of Covariance: ESE I

surprisingly high scores for control students. In contrast to the ESE I control group, the pattern of the ESE II control-group mean scores does not show any startling gains from one test to another but only the steady increase from the test-retest effect.

	Mean		S.D. (w/in)	Regression Coefficient between Pre- and Post-	S.E. (Beta)
	Program	Control			
Pretest	12.95	12.00	3.60	—	—
Initial Post-test	19.62	12.61	3.88	.57	.06
Second Pretest	17.45	13.58	4.08	.75	.06
Final Post-test	18.22	14.22	4.24	.74	.06

Table 5. Means, Regression Coefficients, and Standard Errors: ESE II (N = 223)

A possible explanation of this unexpected ESE I control-group test performance is the radical psychological change in students during fourth grade. According to Piaget's theory of cognitive development,⁵ children around the age of eight to ten years (fourth graders) shift their cognitive perspective from egocentricity to a greater capacity to "take the view of the other." In

other words, children of this age cease to assume that the world and ideas center around themselves and their own activities: they become less subjective and more objective. As part of this development, one could expect fourth graders to be more successful on the objective, multiple-choice tests, in which they have to "take the view of the other," that is, of the person who had written the test.

This kind of development may have been responsible for the fourth-grade control-group gain in spite of the fact that the test was quite technical and, specifically, about economics. Some answers are probably easier to guess than others once the student has realized the necessary attitude or perspective, and it is likely that the high scores of ESE I control students on the first post-test result from their improved skill at guessing answers.

Source	Multivariate F-ratio	df	p (<)
Test Form	1.92	3, 216	.13
Treatment (Eliminating Form)	69.07	3, 216	.00
Form X Treatment	1.86	3, 216	.14

Table 6. Analysis of Covariance on Mean Vectors: ESE II

In light of this important psychological change in fourth graders, we should recon-

⁵ John Flavell. *The Developmental Psychology of Jean Piaget*. Princeton, New Jersey: D. Van Nostrand Company, Inc., 1963; and J. McVicker Hunt. *Intelligence and Experience*. New York: Ronald Press Company, 1961.

Test Form	Initial Post-test	37.34	3.40	1,218	.07	3.40	1,218	.07
	Second Pretest	.14	.02	1,218	.90	1.50	1,217	.22
	Final Post-test	16.53	1.53	1,218	.22	.85	1,216	.36
Treatment (Eliminating Form)	Initial Post-test	2289.72	208.66	1,218	.00	208.66	1,218	.00
	Second Pretest	538.17	57.61	1,218	.00	.03	1,216	.85
	Final Post-test	594.75	54.86	1,218	.00	.20	1,216	.66
Form X Treatment	Initial Post-test	.38	.04	1,218	.85	.04	1,218	.85
	Second Pretest	1.20	.13	1,218	.72	.28	1,217	.60
	Final Post-test	29.27	2.70	1,218	.10	5.27	1,216	.02

Table 7. Multivariate Analysis of Covariance: ESE II

sider our analysis of the ESE I data. We contend that the lack of a statistically significant difference between ESE I experimental- and control-group scores on the second post-test should not be taken to mean that the experimental students had forgotten, to any significant degree, the economics they learned the year before but, rather, that the control students developed a kind of test-taking skill which inflated their scores. (The Step-down F-tests [see Table 4] seem to support this conclusion.)

Summary

The implications of our evaluation of ESE I and ESE II can be summarized in three points:

1. A program of economic studies such as these, using the problematic, social science approach, can be taught effectively as part of the elementary school curriculum.

2. An effective evaluation of a program through a formal and systematic study of what the students learn and how well they remember it may provide a reasonable justification for that program, since the inclusion of an experimental program in a curriculum cannot be justified by theoretical arguments alone.

3. In designing tests for an evaluation study such as the one described above, one must take into consideration the developing cognitive capacities of children. Their developmental level will affect not only the form and types of concepts appropriate for inclusion in the program but also the type of tests most appropriate for evaluating the learning of those concepts. This consideration is especially important when studying retention in an experimental design that uses the same instruments over a period during which significant cognitive changes are occurring in the subjects: one must not confuse psychological development with retention of program content.

References

Donald Campbell and Julian C. Stanley. "Experimental and Quasi-Experimental Designs for Research on Teaching." In: N. L. Gage, editor. *Handbook of Research on Teaching*. Chicago: Rand McNally & Company, 1963. pp. 171-246.

John Flavell. *The Developmental Psychology of Jean Piaget*. Princeton, New Jersey: D. Van Nostrand Company, Inc., 1963.

J. McVicker Hunt. *Intelligence and Experience*. New York: Ronald Press Company, 1961.

Industrial Relations Center. The University of Chicago. *Elementary School Economics I: Readings, Teacher's Guide, and Projects; Elementary School Economics II: Readings, Teacher's Guide, and Projects*. Galien, Michigan: Allied Education Council, 1964.

Task Force on Economic Education. *Economic Education in the Schools*. New York: Committee for Economic Development, 1961. □

Copyright © 1969 by the Association for Supervision and Curriculum Development. All rights reserved.