Queen's Economics Department Working Paper No. 67

AN INVESTIGATION OF THE EFFECIENCY OF INTRODUCTORY ECONOMICS COURSES

R.W. Crowley
Queen's University

D.A. Wilton
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1-1972

AN INVESTIGATION OF THE EFFICIENCY OF INTRODUCTORY ECONOMICS COURSES

Report #1

Test Specification and Some Preliminary Results

R. W. Crowley

D. A. Wilton

Assistant Professors

Queen's University

Discussion Paper No. 67

January, 1972.

Introduction

In most academic circles, inquiring into the value and/or success of education has been considered akin to questioning motherhood. Attempts to measure systematically the relationship between inputs and educational outputs have been vigorously resisted.

Economic research on the "value" of education has centred largely on two issues:

(1) human capital theory and the structure of earnings,[1] and

(2) the contribution of education to economic growth[2]

Recently, there has also been some work on the distribution of the benefits of education by income class.[3] However, the identification and examination of "production functions" remains largely neglected.[4] Elegant theorizing and sophisticated hypothesis testing are "strangely" absent when it comes to the issue of the academic's own productivity. Caveats and casual empiricism, deplored when employed in other contexts, are repeatedly invoked

---

1. See G.S. Becker, Human Capital, New York, National Bureau of Economic Research, 1964.

2. See E.F. Denison, The Sources of Economic Growth in the United States and the Alternatives Before Us, New York, Committee for Economic Development, 1962.

3. See W. Lee Hansen and Burton A. Weisbrod, Benefits, Costs, and Finance of Public Higher Education, Chicago, Markham, 1969.

4. Even where made, such output valuations are often simplistically assumed to be proportional to inputs, i.e., the more sophisticated material incorporated in a course, the greater the value of teaching! One recent exception is S. Bowles, "Towards an Educational Production Function" in W. Lee Hansen, Education, Income and Productivity, New York, National Bureau of Economic Research, 1970.

to evaluate teaching efforts. Of all professions, one might have expected the academic economist to have carefully researched and successfully implemented an efficient learning mix in terms of teaching input and educational output. However, it is only in recent years that economists have even begun to evaluate alternative educational techniques in terms of educational output.[5]

This paper represents the first report of a project designed to examine the "efficiency" of the first year economics course. The first part of the project is to identify and quantify the factors that account for student performance in the introductory course. Secondly, the amount of "new" economic knowledge retained over subsequent years will be examined. The now-classical "Stigler Law" suggests that if one administers a test on current economic problems to college seniors (or persons five years out of college), all students, regardless of previous economics courses, will perform in a similar fashion.[6] A related issue concerns the reinforcement of knowledge obtained through courses after the introductory course. How is retention related to these subsequent courses; does performance on a standard test improve as more economics courses are added to a student's programme? Finally, with learning and retention

---

5. See two volumes edited by K.G. Lumsden, New Developments in the Teaching of Economics, Englewood Cliffs, Prentice Hall, 1967 and Recent Research in Economics Education, Englewood Cliffs, Prentice-Hall, 1970.

6. G.J. Stigler, "Elementary Economic Education", American Economic Review, Papers and Proceedings, May 1963, 657. Preliminary results by Saunders suggest that there are significant lasting effects from introductory economics courses, although such effects are relatively small (a 3.7 gain on a 33 point scale), R Saunders, "The Lasting Effects of Elementary Economics Courses: Some Preliminary Results, American Economic Review, Papers and Proceedings, May 1971, 242-248.

factors carefully analyzed, an attempt will be made to explore techniques for increasing "efficiency" in the introductory economics course. The ultimate question towards which this research is directed is whether there are significant effects from varying the teaching-learning configuration?

This initial report is divided into three sections. In the first part, we discuss the measurement of output and the construction of a suitable test. The second section is devoted to a brief review of factors found important by other researchers in the explanation of student performance in introductory economics. We conclude with a brief examination of data generated in the preliminary test development stage of the project.

Output Measurement and Test Construction

An essential ingredient of rigorous hypothesis testing is quantification of inputs and outputs. Measurement of output is usually obtained through student performance on a standardized objective test.[7] For a variety of reasons, the multiple choice test is generally considered more effective than the true-false test.[8] As Fels points out, a number of considerations are important in the construction of multiple choice tests: test specification,

---

7. The essay question has serious limitations which are generally not recognized by users. Grading is contaminated by extraneous considerations - handwriting, organizational structure and teacher discrimation among students. Because the questions are difficult to grade in a scientific manner, uniform standards of grading are not maintained - especially as graders are influenced in marking subsequent papers by what they have read in the first papers corrected.

8. For example, true-false questions encourage guessing, and since there are only two choices, the options neither present a continuous line of thought nor force the student into clear-cut discriminations. Multiple choice questions allow for an examination of the distribution of the incorrect answers for any one question, thus revealing the nature of the incorrect thinking.

question testing, editing and norming.[9]

Fortunately, the construction of a suitable test has been greatly aided by development of the Test of Understanding in College Economics (hereafter referred to as TUCE). This test was proposed by the Committee on Economic Education of the American Economics Association and carried out by the Joint Council on Economic Education through a "blue ribbon" committee of economists and psychometricians and the Psychological Corporation.[10] The majority of our questions (62% of the final test) are Canadian adaptations of TUCE, supplemented by questions from workbooks accompanying standard introductory textbooks. Thus, a considerable amount of editing and norming has been externally supplied.

Before discussing the criteria used for question selection and the characteristics of those questions selected, we briefly review our testing procedures and test specifications. Initially, a battery of questions was assembled from a variety of sources. From this array, a large group of questions were tested on a relatively small paid group of first year students. On the basis of these results, sixty-six questions were placed on two forty-five question tests (i.e., twenty-four common questions), and one of these tests was administered to each freshman economics class,

---

9. R. Fels, "Multiple Choice Questions in Elementary Education", in K.G. Lumsden, editor, Recent Research in Economics Education, op. cit., 27-43.

10. For a discussion of the purposes and construction of this test, see R. Fels, "A New Test of Understanding in College Economics", American Economic Review, Papers and Proceedings, May 1967, 660-666. We are grateful to the Psychological Corporation for making this test available to us.

Queen's University, at the beginning of the 1970-71 course (hereafter referred to as the pretest).[11] The test was repeated at the end of the academic year (post test), with some students tested under exam conditions and others tested, without announcement, in class. A control group of students was also selected and tested in an identical fashion. In total, 565 students participated in both tests.

Test specifications are usually given in two ways: subject matter and question type. With respect to the first specification, we preferred to have a fairly broad and even coverage of topics rather than a concentration on a few particular topics. Thus, we imposed the constraint that the distribution of questions be approximately;

45 - 50%	microeconomics
35 - 40%	macroeconomics
10 - 15%	international economics

Each of these broad categories includes a wide range of topics. For example, the microeconomics category includes questions on goods markets, factor markets, firm behaviour, consumer behavior, opportunity costs, public goods, etc.

The second specification of the test concerns the type of questions to be included. Our objective was to select questions which probe the basic understanding of economic principles rather

---

11. The first year course at Queen's University is organized into a large number (varying each year from 13 to 18) of autonomous sections, each taught by a different instructor.

than simply to test whether a student can regurgitate the definition of abstract economic terms. In particular, we attempted to include a substantial number of questions which involve economic policy applications. The TUCE format includes an approximately equal number of questions of the following three types.

(1) recognition and understanding

(2) simple application

(3) complex application.[12]

Most of our TUCE adapted questions are taken from the second category.[13] Given our (policy) "application" objective, many of the first category appeared too descriptive or too "memory-work" oriented. Results of a study by Lewis and Dahl[14] indicated that the complex application category of questions showed minimal association with critical thinking skills, whereas the simple application questions had the most significant associations with critical thinking skills.

Given these two constraints, questions were selected on the basis of their power to discriminate. First, questions which all students answer in a similar fashion, either right or wrong, must be eliminated. Also, questions which "good" students get wrong more often than "poor" students are rejected.[15] In other words

12. For discussion of what these categories encompass, see F. Fels, op. cit.

13. Among the final 45 questions selected (Table 1), 62% are from TUCE; 21% are recognition and understanding; 50% are simple application and 29% are complex application.

14. D.R. Lewis and T. Dahl, "The Test of Understanding College Economics and Its Construction Validity", Journal of Economic Education, Spring 1971, 155-166.

15. See below for a discussion of how "good" and "poor" are specified.

the quality of the student and the correct response rate must be positively correlated. Finally, all incorrect answers should be chosen by some students. Wrong responses which are never chosen are inefficient in the sense that testing time is employed in a way which contributes no useful data for discrimating among students.

Since many of our questions are adapted from TUCE it is worth noting the discrimating power of this test as found by other researchers. Results in the Lewis and Dahl[16] study indicated that TUCE was effective in (1) discriminating between "good" and "poor" economics students and (2) incorporating thinking skills while effectively discrimating on other knowledge (i.e., considerable independence exists between TUCE scores and prior ability). In addition, a study by Buckles and Welsh[17] revealed that TUCE was a valid forecaster of future performance in economics courses.

To analyze the relationship between correct response and student quality, we employ a difference of means test. For each question, students are divided into two groups, those giving correct and those giving incorrect responses. The mean value of the "quality" of each group is calculated, and we simply test whether these two means are significantly different. As a measure of student quality, we have two indices available: (1) the student's final grade for the introductory course, and (2) his

---

16. Ibid.

17. S. Buckles and A.L. Welsh, "The Test of Understanding in College Economics and Implications for Further Research", paper presented to the Southern Economics Association, November 1971.

score on the multiple choice test. Given different weighting schemes (i.e., exercises, essays, mid-terms, etc.) for each of the 13 sections of the introductory course, and the varying degrees of "leniency" of the different instructors, we employ the objective post test score on the other 44 questions as the principal index of student quality.[18]

Table 1 presents calculated t-statistics for the difference in average post test score between those students who answered a particular question correctly and those who answered incorrectly (only those questions selected for the final objective test are shown). In general, students who answer a selected question correctly have a significantly higher post test score on the other 44 questions than those students who have an incorrect response (91% of the t-values are significant at the .05 level and 81% are significant at the .01 level). In no instance is there an insignificant difference in means under both sets of "exam" conditions for a selected question. The performance of the rejected questions (not shown) was much less impressive as only 62% passed the .05 level test and 50% passed the .01 level test. Less than half of the rejected questions revealed significant differences in means for both sets of "exam"

---

18. Such an index implicitly assumes that the other 44 questions, as a group, have discrimating power. As pointed out above, initial pretesting and other independent research on TUCE (the major source of our questions) supports such an assumption.

Characteristics of Selected Questions

	Exam Conditions						Non-Exam Conditions					
	Diff. of means	Distribution of Responses	Pre and Post Test Response Patterns				Diff. of means	Distribution of Responses	Pre and post test Response Patterns			
	t-test		WR	RR	WW	RW	t-test		WR	RR	WW	RW
1	2.86	182,70,5,16	.33	.31	.28	.09	4.05	134,98,13,8	.33	.22	.35	.10
2	2.56	167,45,29,28	.29	.32	.29	.10	3.35	143,30,29,30	.29	.29	.29	.13
3	3.34	219,5, 52,2	.27	.50	.13	.10	1.59	194,3,52,5,	.25	.53	.11	.11
4	6.39	185,16,27,43	.28	.37	.25	.10	5.97	154,18,28,49	.26	.34	.29	.11
5	3.68	243,10,8,18	.36	.52	.08	.04	3.11	204,16,6,27	.35	.44	.13	.07
6	2.45	165,32,57,26	.36	.24	.27	.13	3.75	123,27,69,33	.30	.20	.35	.15
7	5.00	106,72,44,44	.36	.03	.54	.07	2.93	88,55,43,46	.32	.05	.55	.08
8	4.84	227,21,5,26	.23	.58	.09	.10	4.63	206,14,4,29	.21	.60	.08	.10
9	3.00	254,4,15,4	.21	.69	.06	.03	2.77	212,10,19,12	.22	.61	.06	.11
10	4.92	203,23,0,50	.29	.45	.17	.08	4.38	169,24,3,56	.22	.44	.25	.09
11	4.99	165,28,67,6	.32	.25	.28	.14	3.56	119,43,71,10	.27	.19	.39	.14
12	2.17	205,17,19,36	.52	.23	.21	.05	4.17	160,20,31,39	.42	.21	.28	.09
13	3.72	149,113,12,6	.46	.07	.41	.05	1.45	62,164,17,8	.19	.05	.67	.09
14	3.10	106,140,16,9	.26	.14	.42	.18	2.98	83,134,17,16	.19	.15	.46	.20
15	1.84	258,2,13,5	.07	.86	.01	.07	2.02	236,1,11,3	.09	.83	.01	.06
16	2.26	235,18,18,7	.21	.63	.07	.09	1.99	215,6,24,7	.25	.60	.06	.08
17	2.76	48,11,5,10	.33	.15	.49	.02	4.36	57,23,42,16	.22	.16	.49	.12
18	2.87	82,3,8,3,	.42	.40	.06	.12	3.46	126,10,12,5	.40	.39	.12	.09
19	1.87	49,21,10,1	.32	.17	.45	.06	2.59	60,53,8,18	.24	.14	.50	.12
20	2.86	83,2,5,8	.32	.52	.11	.05	4.24	123,9,8,16	.33	.46	.16	.05
21	1.66	58,12,11,13	.54	.08	.36	.02	4.07	79,28,32,11	.47	.05	.42	.06
22	1.70	61,17,2,8	.26	.36	.21	.17	2.90	101,16,13,15	.28	.36	.21	.15
23	3.55	30,5,29,28	.26	.06	.65	.02	4.04	36,10,45,62	.18	.04	.72	.05
24	2.58	44,34,2,14	.35	.12	.50	.04	3.99	73,59,0,18	.32	.17	.45	.07
25	2.74	61,4,11,5	.32	.31	.26	.11	3.83	83,7,28,24	.25	.28	.38	.09
26	3.59	69,7,4,10	.52	.17	.21	.10	4.12	95,19,15,18	.41	.23	.26	.09
27	2.51	61,6,2,13	.26	.37	.31	.06	4.02	78,9,29,19	.28	.22	.37	.13
28	5.28	58,5,9,7	.23	.33	.27	.17	4.93	89,15,9,20	.28	.30	.26	.16
29	2.24	53,21,7,4	.26	.24	.35	.15	2.72	78,39,9,9	.34	.15	.37	.14
30	1.52	53,11,14,5	.32	.23	.36	.10	3.07	80,12,13,12	.28	.23	.40	.09
31	3.81	83,11,47,31	.29	.18	.40	.13	4.43	59,6,22,5	.40	.24	.29	.06
32	4.40	82,43,31,20	.28	.19	.44	.10	3.73	37,14,26,18	.29	.06	.44	.21
33	6.06	153,8,17,3	.28	.56	.05	.12	5.42	63,3,23,6	.21	.46	.15	.18
34	3.78	50,33,34,55	.14	.15	.55	.15	1.73	38,16,21,15	.26	.14	.44	.17
35	5.68	120,9,23,24	.39	.30	.20	.11	5.02	57,2,10,19	.33	.29	.27	.10
36	6.18	86,46,28,10	.39	.09	.48	.05	4.11	38,16,18,11	.33	.09	.50	.08
37	2.93	155,7,11,4	.28	.56	.10	.06	4.17	75,2,7,6	.38	.40	.14	.08
38	4.84	130,12,17,19	.46	.30	.20	.04	5.56	52,9,13,16	.38	.19	.27	.15
39	3.64	123,12,4,38	.28	.43	.17	.12	2.17	57,7,4,23	.31	.28	.26	.15
40	5.07	134,17,15,8	.34	.43	.17	.06	5.97	61,7,5,10	.28	.37	.18	.17
41	3.93	142,15,7,12	.20	.61	.10	.10	4.56	69,8,2,11	.28	.44	.10	.18
42	5.17	130,19,9,11	.61	.14	.21	.03	4.20	58,11,5,5	.46	.15	.33	.05
43	5.90	71,26,35,33	.28	.14	.54	.05	3.74	41,5,15,20	.36	.05	.50	.09
44	4.21	156,4,5,13	.28	.59	.08	.05	4.00	72,4,5,5	.35	.44	.12	.10
45	5.98	121,2,23,29	.39	.28	.25	.08	5.71	50,2,13,20	.33	.19	.38	.09

conditions. Final grades for the introductory course produced similar results. No perverse associations occurred in the selected questions and approximately two thirds of these selected questions had significant positive discriminatory power on the basis of this other student quality index.

Students were further divided into four groups for each question on the basis of whether they answered correctly (R) or incorrectly (W) on the pre and post tests. Group WR represents those students who answered incorrectly on the pretest and correctly on the post test. If one assumes "pure guessing" on the part of students, the following "random" pattern would be expected.

	Fraction of Students
WR	.19
RR	.06
WW	.56
RW	.19

Thus, questions which have the following properties would exhibit "perverse" discrimating ability in the sense that performance is inferior to random guessing.

(a) WW $>$ .56

(b) RW $>$ .19

(c) WR + RR $<$ .25

As shown in Table 1, there are only 1.5% of the test cases which reveal such "perversion".

In addition, we subjected these four groups to variance analysis (our criteria again was mean post test scores). Using standard F-tests, 85 of 90 cases showed significant (.05 level) differences in mean scores for the four categories. For pairwise groupings only the WR - RR and WW - RW pairings had insignificant differences in means.[19] This parallels the Lewis and Dahl finding that prior knowledge and ability (i.e. performance on pre test) is not related to performance on the post test.

Finally, Table 1 records the distribution of responses for each selected question. (The correct response is always listed first). With few exceptions, responses are spread over all possible alternatives.

In summary, we have utilized a number of criteria in question selection. Perhaps the most binding constraint was the coverage of many topics. Frequently, we had to reject a question with "good" discriminating power because we already had filled the "quota" on this particular topic, and insert a question on a different topic even though its discriminating power may have been inferior to the rejected question. Nonetheless, there are extremely few instances of "perverse" discriminatory ability and most selected questions show positive discriminatory power under a variety of tests. Finally, as shown in Table 2 our adapted question norms are remarkably close to norms

---

19. Only 20% of these two sets of pairings had significant differences in means. In direct contrast, 82% of the WR - WW and RR - RW cases had significant differences in means.

NORMS FOR SELECTED QUESTIONS

	QUEEN'S NORMS[a]						EXTERNAL NORM[a]
	POST TEST				PRE TEST		POST TEST
	EXAM CONDITIONS		NON-EXAM CONDITIONS				
QUESTION	Sample Size	Norm	Sample Size	Norm	Sample Size	Norm	Source[b]
1	280	0.65	255	0.53	596	.35	Suits .62
2	280	0.60	255	0.56	596	.41	-
3	280	0.78	255	0.76	596	.63	Tuce .77
4	280	0.66	255	0.60	596	.45	Tuce .63
5	280	0.87	255	0.80	596	.54	Tuce .79
6	280	0.59	255	0.48	596	.36	Suits .60
7	280	0.38	255	0.35	596	.12	Tuce .49
8	280	0.81	255	0.81	596	.69	Tuce .73
9	280	0.91	255	0.83	596	.70	Tuce .81
10	280	0.73	255	0.66	596	.53	Suits .57
11	280	0.59	255	0.47	596	.37	-
12	280	0.73	255	0.63	596	.30	Tuce .63
13	280	0.53	255	0.24	596	.13	Tuce .34
14	280	0.38	255	0.33	596	.34	Tuce .45
15	280	0.92	255	0.93	596	.90	Tuce .85
16	280	0.84	255	0.84	596	.68	Tuce .84
17	98	0.49	158	0.36	288	.24	Tuce .32
18	98	0.84	158	0.80	288	.49	Tuce .81
19	98	0.50	158	0.38	288	.23	Tuce .51
20	98	0.85	158	0.78	288	.55	Suits .67
21	98	0.59	158	0.50	288	.12	Tuce .34
22	98	0.62	158	0.64	288	.52	-
23	98	0.31	158	0.23	288	.08	Tuce .39
24	98	0.45	158	0.46	288	.18	Suits .60
25	98	0.62	158	0.53	288	.37	Tuce .58
26	98	0.70	158	0.60	288	.29	-
27	98	0.62	158	0.49	288	.37	Tuce .58
28	98	0.59	158	0.56	288	.46	Tuce .60
29	98	0.54	158	0.49	288	.33	Tuce .60
30	98	0.54	158	0.51	288	.32	Suits .80
31	182	0.46	97	0.61	308	.32	Tuce .49
32	182	0.45	97	0.38	308	.30	Tuce .33
33	182	0.84	97	0.65	308	.63	Tuce .76
34	182	0.27	97	0.39	308	.31	-
35	182	0.66	97	0.59	308	.38	-
36	182	0.47	97	0.39	308	.15	Tuce .45
37	182	0.85	97	0.77	308	.56	Tuce .74
38	182	0.71	97	0.54	308	.34	Suits .69
39	182	0.68	97	0.59	308	.49	-
40	182	0.74	97	0.63	308	.49	Suits .66
41	182	0.78	97	0.71	308	.64	-
42	182	0.71	97	0.60	308	.19	Tuce .46
43	182	0.39	97	0.42	308	.18	Tuce .53
44	182	0.86	97	0.74	308	.60	-
45	182	0.66	97	0.52	308	.34	Tuce .59

[a] The norm is defined in terms of the proportion correct

[b] Sources are Committee on Economic Education, Test of understanding College Economics and Dan B. Suits, Principles of Economics, New York, Harper and Row, 1970

provided externally for the "original" question. (Only data for the questions comprising our final test are presented).

Previous Research

In this section we review the types of factors other researchers have found important in accounting for learning of freshman economics. The most comprehensive work has perhaps been done by Lumsden; and we use one of his studies,[20] based on British data, as a prototype of current research. His basic research strategy entails a standard regression model with each student representing one observation. The dependent variable is performance on a standardized test by students who have just completed a firt year economics course. The various factors which he considered to be important in accounting for learning among these students are the independent, or explanatory, variables. His results, in a general form, are as follows:

1. Initial Understanding of Economics: At the beginning of the year students were given a preliminary test in economics in an attempt to measure their initial level of economic understanding, and, if possible, to isolate this factor from those contributing to learning during the year. The students' initial understanding of economics was found to be a significant factor in accounting for above average performance at the end of the year.

---

20. K.G. Lumsden, "Research in Economics Education", paper presented to the Canadian Economics Association, June 1971.

2. General Aptitude: As one might expect, the intelligence of the student was significant and positively related to performance on the test at the year's end.

3. Personal Characteristics: Neither age nor year at university was found to be significant; however, sex was a significant factor - males appear to have an advantage over females.

4. Type of School Attended: Type of school attended (prior to coming to university) was not found to be a significant factor in accounting for a student's performance on the post test, although it was significant in the initial test to measure the level of economic understanding (referred to above). This result implies that certain types of schools may produce students with different degrees of preparedness but that the advantage (for whatever reason) disappears by the end of the first year.

5. Selected Subjects Taken at School (Prior to University): 'A' level Economics, 'A' level General Studies, 'A' level Mathematics, 'A' level Economic History, and scholarship Economics all had a significant positive effect on student's performance on the preliminary test:[21] however, only 'A' level Economics and 'A' level Mathematics still had a significant effect on

---

21. The 'A' level tests are achievement tests given before entrance to university.

performance by the end of the year. This may suggest that mastery of the logical structure of mathematics is valuable to performance in economics.

6. Field of Specialization: Economic specialists did better on the post test than others. However, there was no significant difference in performance between students who were specializing in economics plus some other subject (accounting, business or commerce, and math science or statistics).

7. Lectures and Tutorials: The larger was the number of students enrolled in a tutorial, the lower was a student's score on the final test; the greater the number of turorials attended, the higher was his score. Lecture size or attendance was not significant.

8. Course Content: The results seemed to suggest that macroeconomics was easier to grasp than microeconomics, and that not enough time was devoted to the teaching of the latter. There was no significant difference among certain well known, and often used, first year economic textbooks on student learning.

9. Lectures and Tutors: Tutor qualifications had no effect. However, anomalously, student scores were found to increase with the status and experience of the lecturer, but decrease with the age of the lecturer!

10. Student Attitudes: How a student evaluated lectures was found not to be important. However, of substantial importance was the manner in which the students viewed

the usefulness of economics, as measured on a five point scale from useless and harmful to essential.

11. Date When Test Given: The closer was the timing of the test to the student's final exam, the better the student did on the standardized test. This implies that some learning takes place in the review, or in the period of intensive study, which precedes an exam; it might further imply that retention may not be great due to the concentrated time period of learning.

Several problems arise in interpreting such results. First, multicollinearity is likely to be a severe problem. For example, one may have difficulty distinguishing initial understanding of economics from general aptitude and intelligence, or possibly from type of school previously attended.

Perhaps more important, psychological factors important in explaining the learning process are not made explicit in the model. In a recent article, Bach[22] suggests that psychological research indicates that a student's motivation is the most important factor controlling the amount of learning that occurs. The greater the motivation to learn, the more likely it is that a person will learn. Reward is considered to be a stronger inducement to learning than fear of punishment. Prompt feedback and information about his performance increases the speed with which a student acquires knowledge. As well, if the student finds the subject matter "relevant", he will absorb it more quickly. An active involvement in the process is also generally considered to

---

22. G.L. Bach, "Student Learning in Basic Economics: An Evaluated Experimental Course", in K.G. Lumsden, ed., New Developments ..., op. cit.

be important in knowledge acquisition. In addition, students tend to retain material which is well organized, "meaningful" and inter-related according to some consistent internal structure. The retention rate increases on "over-learned" material, i.e., material which has been learned and relearned a number of times, or material on which time above the bare minimum required for learning has been spent.

Unfortunately, Bach offers no quantitative test of these "psychological factors". Indeed, they are undoubtedly very difficult to quantify. Consequently, other explanatory variables in some sense may be "proxying" these psychological factors. Problems of interpretation, therefore, may arise in two ways. First, several independent variables may be rough proxies for the same psychological factor, say "motivation", resulting in insignificance for all. On the other hand, a single explanatory variable may be proxying several psychological factors. For example, the significance of tutorial attendance can be interpreted in a variety of ways. Given the insignificance of the lecture attendance variable, one may simply conclude that the "real" learning takes place in tutorial sessions and lectures should be abolished. However, what does tutorial attendance actually measure: student motivation, better feedback, student participation, relevant subject matter, or whatever? Depending on the answer to the latter question, one may conceivably propose a number of alternatives to simply abolishing the lectures.

While Lumsden has done a reasonably effective job in attempting to specify a wide range of variables accounting for

learning, he has failed to take into account the amount of learning of economics which might have taken place over the year due to extra-curricular factors - such as the media, other courses, general outside reading and conversation. To ascertain this, a control group not taking an economics course could have been selected and tested at the beginning and end of the period to determine the "natural" increase, if any, in learning in an educational environment. Steps would have to be taken to ensure that the control group is similar to the group taking the economics course in general attitudes and habits.

In general, however, Lumsden's findings provide a reliable basis for comparison to our own findings.

Some Preliminary Results

Rather than simply discarding the data after selecting questions for the final objective test, we have utilized them to explore a number of the issues raised in the previous section. However, we emphasize that the following analysis is only a preliminary phase of the project and the reader is cautioned for several reasons. First, rejected questions are included in the analysis and thus our output measure is not as precise as will be used in the next phase of the project. Second, the sample sizes are rather small and compartmentalized. For example, we have two different tests administered in three different ways. A related problem concerns the relatively poor sample coverage. Many observations were deleted because (a) the student was absent for either one of the tests or (b) a full set of explanatory variables

was not available for a particular student. Only 64% of the students registered in the introductory course are covered in the following analysis.

The research strategy followed is much the same as Lumsden's, although we attempt to integrate some variance analysis as well. The student score on the post test (out of maximum 45) is the dependent variable in the regression model. Unfortunately, we are more constrained than Lumsden with respect to data availability for explanatory variables, and thus our list of explanatory variables is somewhat limited.[23]

To measure the student's initial understanding of economics, we simply employ the pretest score. Since information of student aptitudes and intelligence is inaccessible, the pretest score should also "pick up" this latter effect as well. Other student characteristics included are sex, faculty or major, and student year. Dummy variables are utilized for the first two characteristics, with unity representing a female (zero a male) and a student's faculty or major (zero in this case representing non-affiliation with a particular faculty or major). As with any set of dummy variables, one category in each set must be omitted as an explanatory variable from the regression. In the following analysis, "males" and non-science students from the Arts and Science faculty are the suppressed categories. Given the relatively small number of observations, the year variable is represented by a value of one for the first year, two for the second year, etc., rather than another set of dummy variables.

---

23. Several new explanatory variables (e.g. student SAT scores), however, will be available for the next phase of the project.

In addition, the enrolment of each student's class is included as a "class size" variable.

Each student also completed a short questionnaire on his attitudes and evaluation of the course and instructor.[24] An analysis of the questionnaire results revealed substantial co-relations between responses to various questions, and thus only the following two questions are employed to measure student motivation and input.

Ability of Instructor to Arouse and Maintain Interest

A Stimulates intellectual curiosity; vigorous discussion.

B

C Adequate interest level and stimulation.

D

E Indifferent; boredom; almost no stimulation.

Amount of Time Spent on Economics 010

A 10% or less of all study time.

B 15% of all study time.

C 20% of all study time.

D 30% of all study time.

E 40% or more of all study time.

Each question was scored on a 1 to 5 scale, in both cases from (a) to (e). Finally, since twenty-one questions are different on each of the two tests, a dummy variable (with a value of zero for Test A and a value of unity for Test B) is included in the event that the degree of difficulty varies between tests.

---

24. Failure to complete this form, or absence the day it was circulated, reduced the set of observations by 127 (approximately 20% of the registered students).

As indicated above, the post test was administered in two ways. One set of classes wrote the test under formal examination conditions while the other classes were tested unexpectedly in a class period near the end of the course. In addition, a small comparable group of students not enrolled in the introductory course was selected for control purposes. Student characteristics for these three groups are presented in Table 3.[25] In the following analysis the three groups will be examined individually and comparatively.

Turning first to the regression results for the exam condition group (see Table 4), all four regressions are significant at the .01 level (critical F-values are in the 2 to 3 range). With respect to individual variables, the constant and pretest score are highly significant (t-statistics are given in parenthesis below estimated coefficients). Unfortunately, it is difficult to draw straight-forward conclusions concerning these two variables since they reflect a number of factors: randomly correct answers (no penalties are imposed for incorrect responses),[26] the omitted category of each set of dummy variables, initial economic knowledge, student intelligence and ability, etc. In terms of the degree of difficulty of the two tests, Test B produced significantly higher marks, but by only about 3%. One

---

25. Since all Commerce students must take introductory economics, the control group is more heavily weighted with Arts and Science students.

26. The guessing factor, approximately eleven correct responses by chance, is presumably a large part of the constant term.

TABLE 3

CHARACTERISTICS OF STUDENTS TESTED

	First Year ECONOMICS STUDENTS		CONTROL GROUP
	EXAM CONDITIONS	NON-EXAM CONDITIONS	
Sample Size	196	212	30
Faculty Distribution			
Commerce	.22	.33	-
Engineering	.34	.14	.30
Sciences	.13	.17	.13
Other	.31	.36	.57
Year Distribution			
First	.50	.75	.97
Second	.32	.15	.03
Third	.14	.09	-
Fourth	.04	.01	-
Male/Female Ratio	3.91	2.92	3.29
Average Pre-Test Score	19.3	18.3	18.6
(Standard Deviation)	(5.0)	(5.2)	(5.0)
Average Post-Test Score	28.6	25.5	20.2
(Standard Deviation)	(5.9)	(6.4)	(5.7)

TABLE 4

REGRESSION RESULTS

VARIABLES	EXAM CONDITIONS				NON-EXAM CONDITIONS				CONTROL GI	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10
Constant	17.04 (5.85)	19.05 (7.88)	17.24 (5.97)	19.14 (7.95)	11.97 (3.72)	10.08 (3.29)	11.54 (3.60)	9.51 (3.12)	1.35 (.41)	1. (.4
Pre test score	.701 (10.87)	.686 (10.81)	.704 (10.98)	.689 (10.94)	.625 (7.95)	.630 (7.96)	.639 (8.16)	.645 (8.19)	.952 (6.29)	.9 (6.4
Test B	1.46 (2.09)	1.44 (2.06)	1.48 (2.12)	1.45 (2.09)	1.21 (1.51)	1.07 (1.33)	1.32 (1.65)	1.18 (1.48)	.74 (.49)	. (.
Arouse Interest	.407 (1.24)	-	.388 (1.19)	-	-.743 (1.82)	-	-.781 (1.91)	-	-	-
Time Spent	.879 (2.36)	.798 (2.17)	.860 (2.32)	.786 (2.15)	.929 (2.42)	1.043 (2.74)	.892 (2.32)	1.009 (2.65)		
Female	-3.34 (4.01)	-3.29 (3.95)	-3.31 (3.99)	-3.27 (3.94)	-.77 (.77)	-.81 (.81)	-.57 (.57)	-.60 (.60)	2.90 (1.59)	2. (1.6
Year	.275 (.56)	.210 (.43)	-	-	1.036 (1.41)	1.123 (1.52)	-	-	.765 (.19)	-
Engineer	-.727 (.64)	-.754 (.67)	-.427 (.43)	-.522 (.52)	-1.552 (.94)	-1.923 (1.17)	-.278 (.20)	-.556 (.40)	.365 (.20)	.3 (.1
Commerce	-1.479 (1.69)	-1.540 (1.76)	-1.513 (1.74)	-1.564 (1.80)	.705 (.69)	.492 (.48)	.739 (.72)	.517 (.51)	-	-
Science major	-2.379 (2.22)	-2.415 (2.25)	-2.336 (2.19)	-2.380 (2.23)	.138 (.12)	.063 (.05)	.674 (.59)	.643 (.56)	-.391 (.18)	-.4 (.2
Class size	-.075 (3.12)	-.079 (3.31)	-.076 (3.19)	-.080 (3.37)	.017 (.47)	.007 (.20)	.024 (.65)	.014 (.38)	-	-
R^2	.482	.480	.484	.482	.266	.258	.262	.253	.564	.5
F-test	19.12	21.01	21.28	23.72	8.64	9.13	9.34	9.92	7.25	9.
Standard Deviation	4.22	4.23	4.22	4.22	5.49	5.53	5.51	5.54	3.78	3.

rather interesting finding is the significance of the class size variable. Such results suggest that an increase in class size by 10 students results in a 2% reduction in post test performance.

The performance of the "questionnaire" explanatory variables is mixed. Increasing amounts of time spent on studying economics, as indicated by the student, are reflected in significantly higher test scores. On the other hand, the instructor's ability to arouse interest in the course had no significant effect on student performance. (The effect that is present is "negative", i.e., the more indifference and boredom, the higher the test score!)

Of the various student characteristics examined, the most striking result concerns the relatively poorer performance of females on the post test. Not only is this variable highly significant, its coefficient is relatively large (the equivalent of approximately 7.5%). As also found by Lumsden, the year variable is insignificant. Finally, science majors have significantly lower scores while engineers exhibit insignificant differences in their post test scores.[27] If one weakens the significance test to the .10 level, the commerce students would join the science majors with lower post test scores.

Since there is often considerable conjecture concerning economics performance by students from various faculties, Table 5 presents regression results for each of the four student

---

27. The structure of the two degree programmes may account for this result. Engineers must include their economics mark in the degree average whereas science majors are not required to (a simple pass is all that is required).

categories separately. Similar sign patterns persist, with most of the variables remaining significant. The calculation of the appropriate F-value reveals that the four student categories are significantly different, perhaps suggesting that the simple intercept dummy variables in Table 4 are inappropriate. However, the ranking by faculty as given by the intercept dummy variables in Table 4 does correspond to the mean post test scores presented in Table 5.

Returning to Table 4, the overall fit is much poorer for the non-exam condition group. While the constant, pretest score, and time spent variables retain significance, all other variables are insignificant. In particular, student sex and faculty variables have standard errors far in excess of coefficient values. On the other hand, the "interest arousal" variable is very close to being significant. Thus, for the group of students who have not prepared for an examination, the more the instructor stimulates interest (response (a)), the higher the post test score.

Results for the control group simply indicate that these students approximately retain their pretest score with all other factors being insignificant.

Turning to performance between groups, we first present results which tests the differences between means. For the pretest scores, there are no significant differences (.05 level) between any pair of the three groups. In other words, statistically they all perform in an identical fashion on the pretest. However,

TABLE 5

REGRESSION COEFFICIENTS FOR
STUDENTS CLASSIFIED BY FACULTY AND MAJOR

(EXAM CONDITIONS)

VARIABLES	COMMERCE	SCIENCE	ENGINEERS	OTHERS
Constant	17.94 (3.50)	5.46 (.69)	15.71 (2.84)	19.14 (3.46)
Prescore	.727 (5.34)	.853 (4.08)	.781 (6.48)	.599 (4.86)
Test B	.064 (.04)	-1.08 (.49)	1.77 (1.64)	2.60 (1.58)
Arouse Interest	-.142 (.20)	.800 (.731)	.723 (1.20)	.226 (.37)
Time Spent	1.483 (2.14)	.106 (.11)	.328 (.33)	1.166 (1.81)
Female	-5.25 (2.96)	-1.43 (.73)	-3.58 (.79)	-2.56 (2.02)
Year	:861 (.48)	2.679 (2.05)	-.147 (.17)	-.206 (.24)
Class Size	-.118 (2.27)	.029 (.48)	-.086 (2.17)	-.088 (1.54)
R^2	.516	.599	.434	.402
F-Test	7.55	6.11	8.23	6.67
Standard Deviation	4.01	3.69	4.32	4.55
Sample Size	44	25	67	60
Mean Post Test Score	27.91	26.08	28.87	29.70

for the post test results, all pairings are significantly different at the .01 level.

Pairings	Difference in Average Post Test Scores	Calculated t-value for the Difference
Exam conditions with non-exam conditions	3.00	4.93
Exam conditions with control group	8.39	7.46
Non-exam conditions with control group	5.39	4.76

The group taking the test under exam conditions significantly outperforms the other two groups on the post test; the largest gain is over the control group. The non-exam conditions group significantly outperforms the control group.

The results are confirmed by F-tests performed on the previous set of regression results presented in Table 4. For example, regression results for the exam conditions group are significantly different than regression results for the other two test conditions (calculated F-values range from 2.4 to 5.3, always exceeding the .01 critical value).

Finally, we compute significance tests for the gain in test score within each group. The two groups of introductory economics students had significant gains of 9.2 and 7.2 over their pretest scores. (Calculated t-values for the differences in means tests were 16.8 and 12.8 for the exam and non-exam groups respectively). While the control group also registered a gain in performance (1.6 questions), this gain is not significant (a calculated t-value of only 1.15).

Conclusions

As stated at the beginning of the last section, these results are only preliminary to the main part of the project. They do, however, suggest a number of interesting conclusions.

(1) While "non-exam condition" introductory economics students significantly improve on pretest scores by attending economics classes, it is very difficult to account for such learning. Only pretest scores, student time input and perhaps "interest arousal" are significant factors. Since the variance for this group is almost twice as large as that for the other group of economics students, varying study habits and subsequently the extent or student preparation for an "unannounced" exam may be the principal explanatory factor which is omitted.

(2) Even though the control group improved marginally in performance over the year, students registered in economics did significantly better. While the gain obtained by the control group is not significant, it does suggest that some "natural" learning may be taking place over the year, and that a small part of the gain registered by economics students is not related to the courses itself.

(3) Finally, the post test performance of the "exam conditions" group is significantly better than the

same group's pretest performance at the start of the year and the post test performance of the other two groups. In other words, exam preparation and course instruction are significant factors. Furthermore, a number of determinants of this significant gain were isolated;

(1) score on the pretest

(2) time spent on economics

(3) sex

(4) faculty or major of the student

(5) class size.

In the next phase of the project, these tentative conclusions will be subjected to much more careful scrutiny. The weeding out of a number of questions which are "weak" in discrimating ability will improve the precision of the output measure. New explanatory variables will be available and others refined on the basis of these preliminary results. Perhaps most important, sample coverage will be greatly improved and the size of the control group will be tripled.