DOCUMENT RESUME

ED 274 676 TM 860 535

AUTHOR Duckworth, Kenneth; And Others
TITLE The Relationship of High School Teachers' Class Testing Practices to Students' Feelings of Efficacy and Efforts to Study.
INSTITUTION Oregon Univ., Eugene. Center for Educational Policy and Management.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Mar 86
NOTE 89p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Academic Achievement; *Classroom Techniques; *Educational Testing; High Schools; High School Students; Locus of Control; Predictor Variables; Questionnaires; Secondary School Teachers; Self Concept; Student Attitudes; Student Evaluation; *Student Motivation; *Study Habits; Teacher Influence; Teacher Made Tests; Teacher Student Relationship; *Teaching Methods
IDENTIFIERS *Efficacy; *Effort

ABSTRACT

The relationship between high school students' feelings of efficacy and efforts to study and teachers' classroom testing practices was examined. Questionnaires were administered in four high schools in biology, geometry, English, and United States history classes; a total of 69 classes participated. Some teachers were also interviewed. Students' efforts to study were indicated by their responses to items on time spent doing homework, completing homework on time, answering questions in class, soliciting teachers' help, listening in class, being late to class, and cutting class. Students' feelings of efficacy versus futility were measured by items about obtaining rewards as a result of effort. Academic motivation was indicated by how important it was to do well and by the desired grade. Results confirmed the expected positive relationships of motivation and efficacy to effort. The predicted negative relationship of futility to effort was weaker. Results of examining teachers' testing practices indicated that students' perceptions of communication, feedback, correspondence, and helpfulness were strongly interrelated; they were also related to efficacy and self-reported effort. The questionnaires are appended. (GDC).

The Relationship of High School Teachers' Class Testing Practices to Students' Feelings of Efficacy and Efforts to Study

by
Kenneth Duckworth, Glen Fielding,
and Joan Shaughnessy

March 1986

Center for Educational Policy and Management
College of Education
University of Oregon
Eugene, Oregon 97403

The preparation of this final report was made possible through an Institutional Grant awarded by the National Institute of Education to the Center for Educational Policy and Management. The opinions expressed in this report do not necessarily reflect the positions or policies of the NIE or the U.S. Department of Education.

Table of Contents

I. Overview of the Study 1
Purpose 1
Research Procedures 4
Sample Characteristics 6

II. Developing a Model of Influences on Students' Efforts to Study . . . 11
Students' Efforts to Study 11
Students' Motivation 13
Students' Feelings of Efficacy versus Futility 14
Students' Academic Aptitude 17
Estimating the Model with Class Means 18
Course Tracking and Effort 18
Summary 19

III. Teachers' Class Testing Practices and Student Efficacy and Effort 20
Student Clarity and Teacher Communication and Feedback 21
Student Perception of Correspondence and Teacher Construction of Tests 27
Student Futility and Teacher Helpfulness 30
Summary 34

IV. Influence of Policy and Collegiality on Class Testing Practices . . 36
School Administrative and Departmental Contexts of Class Testing 37
Collegiality among Teachers of the Same Course 41
Overview of Findings and Implications for Research and Practice 50

References 53

Appendixes 54

The Relationship of High School Teachers' Class Testing Practices to Students' Feelings of Efficacy and Efforts to Study

by
Kenneth Duckworth, Glen Fielding, and Joan Shaughnessy

I. Overview of the Study

Purpose

This paper develops and tests a model of the linkage between high school students' feelings of efficacy and efforts to study and high school teachers' class testing practices. Data used to test the model include responses by teachers and students to questionnaires administered in four high schools in four types of classes: biology, geometry, English, and U.S. history. Data were collected in late November, 1984. The primary analytic methods are comparison and correlation.

With the recent wave of criticism of high schools and the growing demand that schools should expect higher levels of academic achievement from students, the identification of points of leverage on students' efforts to study is an important task for researchers. We attach special importance to students' feelings of efficacy, a concept that means for us the belief that effort pays off. If students regard class tests, which contribute significantly to students' cumulative record of academic success or failure, as unpredictable hurdles where the luck of the draw matters as much as students' preparation, then it may be difficult to increase their efforts to study.

These considerations indicate the relevance to secondary schools of a model of academic effort rooted in an expectancy theory of work performance: level of work effort is hypothesized to be a function of valuation of rewards from work and belief that effort will lead to rewards (Lawler 1976). Natriello (1982, 1983) has applied such a theory to secondary school student evaluation processes and their effect on student disengagement. Natriello found that student-perceived "incompatibilites" in the evaluation system, including unpredictable evaluations, were associated with student reports of disengagement, including withdrawal from efforts to succeed. These fundamental assumptions are tested in Section II of this report, which examines correlations of high school students' efforts to study with their motivation to do well in a class and their feeling of efficacy--belief that studying will enable them to do well on class tests. Parallel analyses at both individual and class levels are conducted for biology, geometry, English, and U.S. history classes. The analyses used to test the model take into account differences in students' academic ability.

Section III focuses on the relationship of students' feelings of efficacy to their perceptions of class testing, particularly the predictability of test coverage. In addition to exploring relationships among students' perceptions and feelings, Section III relates students perceptions and feelings to specific teaching practices. Hence we attempt to trace student perceptions of teaching practices to actual practices by comparing class means on these variables to variation in teaching practice as reported by teachers on a questionnaire. The variables in teaching practice studied are derived from general principles of mastery learning as elaborated by Fielding in a rationale for integrating teaching and testing (Fielding and Schalock 1985).

Among the teaching practices found to be effective in elementary school classrooms are teachers' integration of teaching and testing. Fisher and his colleagues (1980) found that students' academic learning time was positively related to working on tasks that were adapted to their learning level and that focused on knowledge and skills that were to be tested. This finding has encouraged advocates of mastery learning programs to develop programs for training teachers, including high school teachers (Fitzpatrick and Charters 1986), to rethink the "forced march" approach to curriculum coverage and to make assessment of student achievement an integral part of instruction rather than a detached mode of compliance with school grading requirements. Fielding and Schalock (1985) have identified key teacher practices in class testing, including formulation and communication of expectations, purposive test construction, and follow-through on test results by adapting teaching for students whose performance indicates inadequate learning. The present study extends the investigation of class testing practices and tests three hypotheses about their effects on students' feelings of efficacy.

The descriptive-correlational analyses of class testing practices reported in this paper were stimulated by an experimental study of a staff development intervention to improve those practices (Fielding, Shaughnessy, and Duckworth 1986). The research reported in this paper was conducted simultaneously with that experimental study, which prompted us to study the school context of high school teachers' class testing practices--especially school and departmental policies and collegial faculty relationships that might strengthen such practices. Studies of effective schools have emphasized the importance of instructional leadership and a collegial climate (Purkey and Smith 1985). Also, a study of elementary school teachers' use of some of the assessment practices advocated by mastery learning found that those practices were strengthened by supportive collegial interaction among teachers (Intili 1977). In Section IV, we report evidence from the present study about the relationship of testing practices to policy and collegiality.

Research Procedures

The design of the research was to compare and correlate teacher practices and student perceptions in high school classes. We suspected that differences in subject matter structure would affect testing practices and their effects on students, so to ensure that the model we were developing would be generalizable, we decided to develop it simultaneously in four different subjects. Twenty classes per subject was set as the sampling goal. In order to obtain a representative sample of classes for each subject, we decided to study a cross-section of all the classes in a small number of high schools rather than recruit isolated and possibly exceptional teachers from a large number of schools. Therefore, we sought five classes each in subject in each of four high schools.

In studying teacher and student perceptions in a large number of classes, we deemed questionnaires the most efficient source of data. Early contacts with administrators in potential research sites indicated reluctance to commit too much school time to questionnaire administration, so it was decided to focus on one course in each subject and to administer questionnaires to five teachers of that course in each high school. In order to gather information about school and departmental policies on class testing, we interviewed administrators and chairpersons of each participating department during the spring of 1984. At that time, we obtained their cooperation in selecting teachers and classes to receive the questionnaire in November, 1984. Each of those teachers was instructed to choose one class, or "section" of the course, to fill out the student questionnaire.

Selection of classes in which to administer the questionnaires was based on our desire to study a cross-section of the students taking a particular subject in each school. Where all classes in a subject were in principle heterogeneous, we asked each teacher of the subject to select one class. Specific sampling decisions are described under "Sample Characteristics" below. Where sections were tracked, we asked teachers of each track to select one class. (We excluded upper-level "advanced placement" classes from our study, however.) Although we sought five teachers per course, it transpired that in some schools, some courses were taught by fewer than five teachers. Furthermore, one English teacher withdrew from the study upon receiving the questionnaires. Thus the actual sample included fewer teachers and classes than planned: 16 for biology; 18 for geometry; 18 for English; and 17 for U.S. history.

This model development study was linked with a concurrent experimental test of a staff development intervention in science courses (Fielding, Shaughnessy, and Duckworth 1986). Thus in two of the biology classes in each school, the questionnaires followed a series of interviews with teachers about testing practices and concomitant observations of class sessions when tests were passed back and discussed. Moreover, in two of the four schools those teachers also expected to participate in a series of workshops regarding class testing practices.

The questionnaires were pilot tested in spring 1984 with five science teachers in a nonparticipating district. We eliminated items with insufficient variation and items where correlations did not indicate sufficient construct validity. The final questionnaires are included in Appendix A. Both teacher and student questionnaires asked respondents to answer most of the questions in terms of the specific class during which the questionnaires were administered.

Sample Characteristics

The four high schools participating in the study were located in two districts in the Northwest. Each district had a reputation for academic excellence, supported by the fact that its students' Scholastic Aptitude Test scores were higher than the state average. District A was located in a medium-sized city. The city housed one of the state universities, and the population ranged from upper-middle-class professionals to working-class people. Schools 1 and 2, both in District A, were schools including grades 9-12 and enrolling 1,000-1,300 students each. District B was a suburban district near a large city in the same state as District A; it served a largely middle-class and upper-middle-class clientele. Schools 3 and 4, both in District B, were larger schools of 1,500 or more students each, serving grades 10-12. Whereas District A was on a trimester system, District B was on a quarter system.

Administrative Characteristics. In all four schools, the primary responsibility for administering the instructional program fell to an assistant principal or vice-principal for curriculum, although the principal remained formally responsible and would always be informed about developments. In the organizational structure of all four schools, the person below the assistant principal for curriculum was the department chair. Department chairs were largely administrative posts rather than academic leadership posts, although this varied by school, to be described in Section IV. All four schools used the standard letter grading system and mailed academic progress reports or warning notices to parents midway in the quarter or trimester. School 1 was unique among these schools in having an eight-period day, during which students had two free periods. This allowed for more interaction between teachers and students in remedying student learning problems than was possible in the other three schools, where students were in class each period during the regular school day.

Curricular Characteristics. As mentioned above, we needed to include several sections of each subject in each high school. Our plan was to focus on courses taken by a cross-section of students in each school, which dictated that we seek courses taken in the sophomore year (District B lacking the ninth grade) while curricular branching was at its lowest level. Specific sampling decisions regarding classes to be studied in each subject were made within a framework of the school's curricular offerings. We discuss the sequence of courses in each subject that informed our sampling decisions.

In science, each of the four schools offered a standard curriculum: biology for sophomores and chemistry, physics, and advanced biology for juniors and seniors who wanted a fuller college preparation. The biology course was the only one of the three subjects to approximate a "required" subject, although it was possible for students to satisfy high school graduation requirements without taking biology. The schools varied in their curricular differentiation with regard to academically weak students and academically strong students. Schools 3 and 4 had more offerings for advanced students, whereas Schools 1 and 2 had more offerings for students with low skill levels.

In mathematics, as in science, all four schools offered the standard sequence of college preparatory courses--algebra in the 9th grade (taken in the junior high schools of District B), geometry in the 10th grade, advanced algebra in the 11th grade, and precalculus or calculus in the 12th grade. This regular track of courses was the stem for numerous branching, however. Some students would never take algebra but would take courses such as "math for life" or "intermediate math" to complete their high school graduation requirements. Other students would be shunted after algebra into a series of "elements of . . ." or survey courses that covered the standard college preparatory curriculum in less depth and with less speed. For example, in Schools 3 and 4, students might take "survey of math" in the 10th grade, which included both geometry and algebra. Still other students, especially in District B, might be shunted into accelerated courses after algebra and would complete the equivalent of a college course in calculus in their senior year.

The English departments in the four schools were alike in offering one standard course for sophomores. School 3 offered several sections of advanced sophomore English; School 4 did not, although the more able student could petition to waive sophomore English and proceed directly into literature and composition courses usually reserved for the upper grades. School 1 offered both "survey" and "advanced" sophomore English in addition to regular sections; School 2 offered only "advanced" and regular sections. In all schools, there were remedial courses that students might take instead of sophomore English, but these classes enrolled only a few students and were not included in our study.

The social studies departments varied in the sequencing of courses, and there was no single course taken in the 10th grade at all four schools. Because U.S. history was a course required of all students, we chose that course to study. It was generally taken by juniors, although School 2 allowed sophomores to take it. School 3 offered "basic" and "advanced" sections in addition to regular U.S. history; Schools B and C offered only regular sections; School 2 offered "developmental" sections in addition to the regular U.S. history course. We included each type of class offered in our sample.

Figure I-1 shows the actual sample of classes studied. Of the 16

Figure I-1
Configuration of the Sample of Classes: Number of regular, high-track, and low-track classes in each subject in each school.

School/Track		Biology	Geometry	English	U.S. History
	High			1 "Advanced"	
1	Regular	3 Regular	3 Regular	1 Regular	4 Regular
	Low		2 "Elements"	1 "Survey"	
	High	1 "Honors"	1 "Advanced"	1 "Advanced"	
2	Regular	2 Regular	1 Regular	4 Regular	2 Regular
	Low	1 "Skills"	1 "Elements"		1 "Developmental"
	High			2 "Advanced"	1 "Honors"
3	Regular	4 Regular	3 Regular	3 Regular	3 Regular
	Low		2 "Survey"		1 "Basic"
	High				
4	Regular	5 Regular	3 Regular	5 Regular	5 Regular
	Low		2 "Survey"		
	High	1	1	4	1
All	Regular	14	10	13	14
	Low	1	7	1	2
Total		16	18	18	17

biology classes, only 2 were designated as other than regular classes. In School 2, we included one "study skills biology" class and one "honors biology" class along with two regular biology classes.

On the other hand, only about half the mathematics classes included were regular geometry classes. Of the 18 classes included, 7 were lower-track classes: 2 classes of "elements of geometry" in School 1; 1 "elements of geometry" class in School 2; and 2 "survey of math" classes in each of Schools 3 and 4. This higher proportion of lower-track classes is attributable to the fact that geometry is the least widely taken of the four subjects. In order to include a number of teachers roughly equivalent to the numbers included for the other three subjects, we had to extend the sample to courses taken by students weak in math (all of which included geometry subject matter, although the "survey of math" classes also included algebra). The sample also included one "advanced geometry" class for 9th graders in School 2.

With sophomore English, the problem was different. All students took this course, and we included only one lower-track -- "survey" -- section in School 1. However, three of the four schools offered several sections of "advanced" sophomore English, so for representativeness we included four such classes -- one in each of Schools 1 and 2 and two in School 3.

Finally, 14 of the 17 U.S. history classes were also "regular" classes; we included one "developmental" section in School 2 and one "basic" section in School 3--each of which was a small class--and one "honors" section in School 3.

This unevenness in sample constitution, especially in geometry and English, is taken into account in using questionnaire data to develop the model.

<u>Teacher and Student Characteristics.</u> Ideally, we regarded variation

in teacher and student characteristics across subjects as undesirable for our model-building effort. However, with a small sample, some variation was inevitable. Table I-1 shows some information about the teachers and students who participated. There were differences in the length of time teachers had been teaching in their current schools. Of the history teachers, none was in the first or second year at his or her current school, and nearly half had been there more than ten years. Of the English teachers, nearly half were in their first or second year, and less than one fourth had been there more than ten years. Also, more English and geometry teachers than biology and history teachers reported that they had more than three preparations each day. On the other hand, English teachers reported teaching fewer students overall, which implies that their classes were smaller than, for example, history teachers' classes. This in fact was sometimes a matter of policy when the English class involved considerable student writing and therefore more teacher time to grade papers.

The student data shows far more similarity across subjects, except for the expected difference in proportion of students in the 10th grade or below. The immediate impression is that these courses in these schools enrolled educationally successful and ambitious students. From 63 to 75 percent of the students in each subject planned to go to a four-year college, more than half reported that their main courses were college-preparatory subjects, and more than half reported that at least one parent had graduated from college. Furthermore, when asked about their previous year's grade point average, from 68 to 81 percent replied that it was in the "A" or "B" range. In general, the geometry students seemed the most academically oriented, borne out also by the finding that nearly a third of them indicated that they were taking the class in which the questionnaire was administered mainly because they wanted to rather than because it was required or

Table I-1
Characteristics of Teachers and Students in Sample
(Percentages of teachers and students, in each of four subjects and in all, selecting questionnaire responses)

Characteristic	Response	Biol	Geom	Engl	Hist	All
Teachers:						
Years in school	More than 10	38	41	22	45	37
	1 or 2	25	29	44	0	25
Number of preps each day	3 or more	8	53	66	29	40
Number of students taught	More than 140	44	29	11	53	39
Students:						
Year in school	9th or 10th grade	88	62	97	14	65
Plans after high school	4-year college	69	75	67	63	69
Main classes	College-prep	63	60	63	59	61
Grade point average	A or B	77	81	77	68	76
Parents' education	College graduate	59	66	63	56	61
	Attended college	21	20	17	22	20
	High school only	20	14	19	22	19
Main reason for taking this class	I wanted to	21	31	14	11	20
	It was required	79	69	86	89	80
Number of teachers in sample*		16	18	18	17	69
Number of students in sample*		391	478	420	405	1694

* Actual numbers for each item depend on numbers of missing responses.

recommended. (All four of the subjects we studied were in fact required or recommended for most students at these high schools.)

II. Developing a Model of Influences on Students' Efforts to Study

In this section, we present the components of a model of influences on students' efforts to study. Using an expectancy theory of work performance (Lawler 1976; Natriello 1982, 1983), we treat such effort as a positive function of the value of work rewards to the student and of the student's expectation that effort will obtain rewards. We begin by using data from the student questionnaire to create an index of efforts to study, and we explore the relationship of this index to other measures of students' academic effort and participation in class. Then we investigate influences of students' motivation and students' feelings of efficacy on their efforts to study. Next, we control on the effects of differential student academic aptitude on variables in the model.

Initially, we develop the model at the individual level in each subject as well as in the four subjects taken together. In exploring effects of academic aptitude, however, we augment individual-level analyses with class-level analyses and control on differences in track level among classes in each subject. Development of the model at the class level establishes the foundation for the analyses in Section III, which use data from both student and teacher questionnaires to identify the effects of class testing practices on students' feelings of efficacy and efforts to study.

Students' Efforts to Study

The student's effort to study is the primary criterion variable of the model. We focus primarily on the subjective sense of expending effort in study and in classwork in general, although we examine relationships of this subjective sense of effort to various other measures of academic effort and class participation.

Our criterion variable was measured by student responses to two questionnaire items, which asked whether the student "nearly always," "often," "occasionally," or "almost never" studied the material thoroughly before tests in the class in which the questionnaire was administered (item 31) and worked hard in that class (item 16). Table II-1 shows the distribution of student responses to these items. Approximately one fourth of the students reported nearly always studying thoroughly before tests. Approximately one third of the students reported that they nearly always worked hard in the class.

The correlation between studying thoroughly and working hard ranged from .39 to .52, as is shown in Table II-2. The two items were averaged into a single index of Effort. Means on Effort are shown in Table II-1.

Because the present analysis of influences on students' efforts to study is part of a broader investigation of students' academic effort and participation in school, it was important to us to establish that this subjective sense of effort was related to other measures of students' academic work and participation in class. Hence we computed the correlations of the Effort index with student responses to each of seven questionnaire items measuring work and participation. One item asked how many hours the student spent each week doing homework or studying for the class. The other six asked how often the student:

had homework done on time
asked questions in class
solicited teacher help after class
"tuned out" during class discussions
was late to the class
cut the class.

Distributions of student responses to these items are included in Appendix B,

Table II-1
Students' Effort, Motivation, Efficacy, and Futility
(Percentage of students, in each of four subjects and in all, selecting questionnaire responses)

Questionnaire Item	Response	Biol	Geom	Engl	Hist	All
31. Before each test in this class, I study the material thoroughly	Nearly always	23	18	30	28	24
	Often	35	32	33	35	34
	Occasionally	31	32	28	28	30
	Almost never	11	17	10	10	12
16. I work hard in this class	Nearly always	29	41	38	29	35
	Often	38	37	38	38	38
	Occasionally	30	19	21	27	24
	Almost never	4	2	4	6	4
10. To me, doing well in this class is	Very important	55	67	63	55	60
	Somewhat important	43	31	33	42	37
	Not important	2	2	4	3	3
8. The lowest grade I would be satisfied with in this course *	A	14	10	16	10	12
	B	40	46	45	35	42
	C	36	42	34	44	39
	D	9	3	6	11	7
32. If I study hard for this class, the effort is rewarded	Nearly always	42	41	47	44	43
	Often	31	29	29	26	29
	Occasionally	20	22	18	20	20
	Almost never	7	8	7	9	8
35. I can predict how well I will do [illegible] a test in this class [illegible] how hard I have st[illegible]	Nearly always	30	25	31	31	29
	Often	29	31	28	32	30
	Occasionally	26	27	28	24	26
	Almost never	15	18	13	13	15
38. Doing well on te[illegible] in this class depends on factors beyo[illegible] my control	Nearly always	7	5	8	9	7
	Often	17	11	14	16	14
	Occasionally	32	31	31	30	31
	Almost never	44	52	48	45	48
11. No matter [illegible] [illegible]rd I work in this clas[illegible] I can't do as well as I would like	Stro[illegible] agree	11	12	11	12	11
	Agree	26	28	26	26	26
	Disagree	46	43	45	45	45
	Strongly disagree	16	18	19	16	18
Indices:						
Effort (avg. items 31+16)	Mean	2.81	2.84	2.96	2.87	2.87
Motivation (avg. items 10+8)	Mean	2.56	2.63	2.65	2.48	2.59
Efficacy (avg. items 32+35)	Mean	2.91	2.82	2.97	2.93	2.90
Futility (avg. items 38+11)	Mean	2.06	2.00	2.05	2.11	2.06
Number of students **		391	478	420	405	1694

* "Don't know" responses treated as missing responses
** Actual number for each item depends on number of missing responses.

Table II-2
Correlations of Effort Index with Other Measures of Students' Academic Effort.
(Individual-level correlations in each subject and in all*)

Items Correlated	Biol	Geom	Engl	Hist	All
Effort index items:					
Study thoroughly x Work hard	.42	.39	.52	.51	.44
Effort index x:					
Hours spent in study and homework	.40	.27	.25	.38	.32
Homework done on time	.52	.41	.62	.52	.51
Ask questions in class	.33	.35	.39	.27	.34
Seek teacher's help	.30	.31	.35	.35	.32
Mind wanders in class	-.34	-.35	-.34	-.39	-.36
Late to class	-.24	-.13	-.37	-.18	-.22
Cut class	-.24	-.14	-.32	-.22	-.22
Numbers of students **	391	478	420	405	1694

* All correlations statistically significant at $p<=.05$ level.
** Actual number for each item depends on number of missing responses.

Table 1. Our concern was their correlations with Effort, which are shown in Table II-2.

Effort exhibited statistically significant positive correlations with hours spent in study or on homework, having homework done on time, asking questions, and soliciting teachers' help. The correlations with having homework done on time were particularly strong. Effort also exhibited statistically significant negative correlations with "tuning out" in class, being late to class, and cutting class, although correlations with the latter two variables were weak, possibly because of the limited variation in response to those questions, which we attribute to students' reluctance (especially in District B) to admit to rule-breaking behavior. Despite these low correlations, however, the general pattern evident in Table II-2 encourages us to regard the Effort index as an indicator of academic effort in general as well as of efforts to study in particular.

Students' Motivation

The first step in our model development was to test our hypothesis that effort to study is a positive function of academic motivation. To measure motivation, we employed two questionnaire items: how important it was to the student to do well in the class (item 10) and the lowest grade in the class that would satisfy the student (item 8). Distributions of student responses to these items are shown in Table II-1. The majority of students in each subject reported that doing well was "very" important. Similarly the majority of students in each subject except history reported that the lowest grade that would satisfy them in the class would be a B or better. The correlations between these two items, shown in Table II-3, were strong enough to warrant averaging them into an index called Motivation. Means on this index are shown in Table II-1.

Table II-3
Correlations of Effort, Motivation, Efficacy, and Futility
(Individual-level correlations in each subject and in all*)

Items Correlated		Biol	Geom	Engl	Hist	All
Motivation index items:						
Lowest grade x Do well		.37	.27	.39	.35	.35
Efficacy index items:						
Effort rewarded x Can predict		.41	.43	.44	.46	.43
Futility index items:						
No control x Can't do well		.33	.29	.25	.28	.29
Effort x:	Motivation	.41	.33	.47	.40	.40
	Efficacy	.28	.27	.49	.25	.31
	Futility	-.06~	.05~	-.23	-.08~	-.09
Motivation x:	Efficacy	.28	.21	.29	.23	.25
	Futility	-.33	-.30	-.27	-.32	-.31
Efficacy x	Futility	-.39	-.28	-.40	-.44	-.37
First-order partial correlation coefficients, controlling on Motivation:						
Effort x	Efficacy	.19	.22	.42	.18	.25
Number of students **		391	478	420	405	1694

* All correlations except those marked "~" are significant at p<=.05.
** Actual number for each correlation depends on number of missing responses.

The functional relationship of student effort to student motivation was operationalized as the correlation of Effort with Motivation. As shown in Table II-3, correlations in each subject were positive and statistically significant. The hypothesis was confirmed.

Students' Feelings of Efficacy versus Futility

We also hypothesized that Effort is a positive function of students' feelings of efficacy in obtaining rewards as a result of effort. The model predicted that efficacy would have a positive effect on effort, because the expectancy theory of work motivation predicts effort as a function of subjective probability of success.

To measure feelings of efficacy, we included four items in the student questionnaire. Of these four items, two were positive measures of efficacy and two were negative measures. The two positive items asked how often the student felt that "If I study hard for this class, the effort is rewarded" (item 32) and that "I can predict how well I will do on a test as a result of how hard I have studied" (item 35). Distribution of responses is shown in Table II-1. The data show that about three fourths of the students in each subject felt that the effort involved in studying hard for class tests was often or nearly always rewarded. However, only about 60 percent of the students in each subject felt that they often or nearly always could predict how well they would do on a test based on how hard they had studied.

The two negative measures asked how often the student felt that "Doing well on tests in this class depends on factors outside my control" (item 38) and how strongly the student agreed or disagreed that "No matter how hard I work in this class, I can't do as well as I would like" (item 11). As the distribution of responses in Table II-1 shows, only a fourth at most of the students in each subject felt that doing well on tests often or nearly always depended on factors beyond their control. A larger proportion of students--about 38 percent--agreed or strongly agreed that they couldn't do as well as they would like regardless of how hard they worked.

Initial analyses of item relationships revealed that the negative measures exhibited a somewhat different pattern of correlations than the positive measures. As we reflected on the wording of these negative items, we sensed that they evoked feelings of "futility" more general than the positive items' specific reference to the rewards for studying for tests. The "futility" items probably tap some more long-lasting attitudes and feelings than the positive "efficacy" items. Hence we decided to distinguish between positive feelings of efficacy about the payoff of studying and negative feelings of futility about succeeding as a result of effort; we expected the latter to more somewhat more general to the personality of the student and the former to be more responsive to situational differences.

Table II-3 shows the correlations of the two positive items and of the two negative items. The two positive items were averaged into an index called Efficacy and the two negative items into an index called Futility. Table II-3 also shows the correlations between the two indices. They were strongly negatively related, as we expected from the correlations among individual items. Both indices are employed in testing hypotheses about the predicted influence of feelings of efficacy on efforts to study (and, in Section III, in analyses of predicted influences <u>on</u> feelings of efficacy).

Table II-3 also shows the correlations of Efficacy and Futility to Effort. Efficacy exhibited moderate positive and statistically significant correlations with Effort. In contrast, the predicted negative relationship between Futility and Effort was statistically significant only for students in English classes and for the whole sample, and even there it was weaker than any of the correlations of Efficacy with Effort. In geometry, the sign of the correlation between Futility and Effort was even positive, although it was insignificant. These findings are consistent with the expectation that feelings of efficacy are significant positive influences on efforts to study; they do not in general, however, support the hypothesis that feelings of futility are significant negative influences on efforts to study. We will retain the Futility index in our analyses as a criterion variable interesting in its own right, but we cease to expect it to mediate the influence of class testing practices on Effort.

Finally, Table II-3 shows the correlations of Efficacy and Futility with Motivation. The findings show that the correlations of Efficacy with Motivation were positive and statistically significant, but they were never stronger than .29. The correlations of Futility with Motivation were negative and statistically significant and generally a little stronger than the correlations of Efficacy with Motivation. Apparently students who are motivated to do well also are somewhat more likely than other students to believe that they can do well.

In order to make doubly sure that the positive relationship between Efficacy and Effort, which is a cornerstone for later analyses, was genuine and not a spurious result stemming from the confounding influence of Motivation, we computed the first-order partial correlation between Efficacy and Effort, with Motivation controlled. (Since Futility was not generally a significant correlate of Effort, we omit that index from this analysis.) The results, shown in Table II-3, confirm in general that the partial correlation coefficients, while smaller than the zero-order coefficients, remain positive and statistically significant. This analysis also supports the hypothesis that Efficacy has a positive influence on Effort.

Students' Academic Aptitude

It was important to us to ensure that the relationships we observed were not attributable to the effect of differences in students' academic aptitude. Perhaps students with a long history of success in school have developed reinforced tendencies to work hard and thus have developed more effective and efficient study habits. Furthermore, students who have succeeded in past courses are likely to feel rewarded by study. Conversely, students who have not done well in the past may feel more hopeless about doing well. Hence for this reason also, the positive relationship of effort and efficacy might be spurious. The sole questionnaire item measuring aptitude was a question asking students what their grade point average (GPA) was during the preceding year. This item was mentioned earlier in describing the characteristics of students in the sample. GPA is not a pure measure of aptitude; it is, among other things, also a measure of effort in past years. Given the likelihood that aptitude and the habit of effort will influence the variables in the model under development, we include it here. Distributions of student responses to the question about GPA are shown in Table II-4. As already mentioned in connection with Table I-1, large majorities of the students in each subject reported having compiled an "A" or "B" GPA during the preceding year.

Correlations of GPA with model variables are also shown in Table II-4. GPA is most strongly related to Motivation. Evidently, students who have done well in the past expect to keep doing well. The correlations of GPA with Effort are also moderately strong. Again, past study habits presumably carry over to the present. However, GPA is not related to Efficacy (except very weakly in English) and only weakly related to Futility. Controlling on GPA in addition to Motivation does little to change the relationship of Efficacy to Effort. The second-order partial correlation coefficients remain statistically significant.

Table II-4
Students' Grade Point Average and Its Correlates
(Percentages of students selecting questionnaire responses, and individual-level correlations, in each subject and in all)

Item	Response	Biol	Geom	Engl	Hist	All
Distribution of responses:						
What was your grade point average last year? (6) **	3.6-4.0 (A- to A)	27	27	30	12	24
	2.6-3.5 (B- to B+)	50	54	47	56	52
	1.6-2.5 (C- to C+)	20	18	18	29	21
	Below 1.6 (D+ or below)	3	1	5	2	3
Correlation coefficients:						
Zero-order coefficients:						
GPA x: Effort		.33*	.20*	.41*	.22*	.29*
Motivation		.54*	.39*	.58*	.49*	.51*
Efficacy		.05	.04	.14*	.05	.07*
Futility		-.20*	-.22*	-.30*	-.21*	-.23*
First-order partial correlation coefficients, controlling on GPA:						
Effort x Efficacy		.28*	.27*	.48*	.24*	.32*
Second-order partial correlation coefficients, controlling on GPA and Motivation:						
Effort x Efficacy		.21*	.22*	.44*	.19*	.26*
Number of students ***		391	478	420	405	1694

* $p <= .05$
** "Don't know" responses treated as missing responses.
*** Actual number for each item and correlations depends on number of missing responses.

Estimating the Model with Class Means

Thus far we have concentrated on analyses at the individual student level. Because the model ultimately aims to relate student variables to teacher practices that largely affect all members of a class, and because class stratification results in ecological segregation of student aptitude and motivation, we now replicate and expand the above analyses with class means of student measures.

In Table II-5 we show the correlations among class means on Effort, Motivation, Efficacy, Futility, and GPA. The class-level relationships are usually as strong as, and often stronger than, the individual-level relationships shown in Tables III-3 and III-4. The small number of classes involved in correlations for each subject, however, make statistical significance more difficult to obtain.

Course Tracking and Effort

As described in Section I, the classes studied included some identified as high-track or low-track classes. It is important to correct for such uneven stratification of courses in class-level analyses, because some of the observed relationship among variables may be attributable to the segregation of more and less able students by track. We created an index called Track by assigning a value of 1 to low-track classes, 2 to regular classes, and 3 to high-track classes. Table II-5 shows the correlation of Track to model variables for each subject. Track was moderately related to Effort, strongly related to Motivation, unrelated to Efficacy, moderately related to Futility, and very strongly related to GPA.

In order to remove the influence of tracking, we recomputed the correlations for the subsample of regular classes (Track=2). These

Table II-5
Estimating the Model with Class Means of Student Responses
(Class-level correlations in each subject and in all, for all classes and for "regular" classes alone)

Indices Correlated		Biol	Geom	Engl	Hist	All
All Classes:						
Track x:	Effort	.38	.40	.46	.41	.34*
	Motivation	.47*	.79*	.90*	.15	.65*
	Efficacy	-.13	.29	-.27	.15	.04
	Futility	-.23	-.59*	-.45*	-.02	-.39*
	GPA	.90*	.92*	.86*	.86*	.73*
Effort x:	Motivation	.42*	.63*	.56*	-.05	.45*
	Efficacy	.33	.67*	.42*	.21	.43*
	Futility	-.22	-.62*	-.61*	.02	-.36*
	GPA	.36	.30	.43*	-.06	.27*
Motivation x:	Efficacy	.27	.48*	-.06	-.06	.15
	Futility	-.56*	-.76*	-.52*	-.52*	-.59*
	GPA	.64*	.66*	.93*	.69*	.78*
Efficacy x:	Futility	-.70*	-.72*	-.57*	-.56*	-.59*
	GPA	-.19	.17	-.15	-.29	-.12
Futility x	GPA	-.18	-.45*	-.47*	-.33	-.39*
"Regular" Classes (Track=2):						
Effort x:	Motivation	.37	.01	.18	.13	.23*
	Efficacy	.49*	.42	.33	.17	.35*
	Futility	-.27	-.15	-.35	.01	-.22
	GPA	.09	.17	.00	.18	.09
Motivation x:	Efficacy	.37	.08	.30	.71*	.18
	Futility	-.52*	-.54*	-.34	-.72*	-.55*
	GPA	.58*	-.14	.74*	.02	.68*
Efficacy x:	Futility	-.27	-.43	-.89*	-.70*	-.60*
	GPA	.09	.28	.17	-.15	-.09
Futility x	GPA	.08	-.10	-.11	.19	-.18
Number of classes (total)		16	18	18	17	69
Number of regular classes		14	10	13	14	41

* p<=.05

correlations are also shown in Table II-5. What is apparent in this analysis is that the relationship of Motivation to Effort almost disappears in this subsample. Only the whole-sample correlation is significant. In contrast, the correlations of Efficacy with Effort remain significant at the whole-sample level, are still appreciable in geometry and English, and even increase above the level of statistical significance in biology. Thus this relationship seems more robust than the relationship of Motivation to Effort.

Finally, we recomputed the correlations for low-track classes in the one subject (geometry) with enough low-track classes (7) to permit correlational analyses. Here the correlations with Effort were as follows:

Motivation	.86
Efficacy	.84
Futility	-.82

All three were statistically significant. Evidently the basic model holds for this sample of low-track classes.

Summary

We have introduced indices of Effort, Motivation, Efficacy, and Futility in developing a basic model of predictors of students' efforts to study. Using individual-level analyses, we found that the expected positive relationships of Motivation and Efficacy to Effort were statistically significant but that the expected negative relationship of Futility to Effort was weaker and was significant only in English and in the sample as a whole. Controlling on Motivation and GPA did little to diminish the positive relationship between Efficacy and Effort.

Then we reanalyzed the model using class means of student data. In general, the relationships were stronger. When we restricted analyses to regular, untracked classes, however, the relationship of Motivation to Effort was noticeably diminished. Efficacy, however, still exhibited a positive relationship to Effort. These analyses encourage us to treat students' feelings of efficacy as a promising mediating variable between teachers' class testing practices and students' efforts to study.

III. Teachers' Class Testing Practices and Student Efficacy and Effort

Our analyses of effects of efficacy on effort warranted the inference that a student's feeling of efficacy was a positive influence on a student's effort to study. In this section we consider teachers' class testing practices that may increase students' feelings of efficacy and therefore also increase effort.

Based on the rationale of a program for integrating teaching and testing developed by one of the authors (Fielding and Schalock 1985), we test three hypotheses. First, we hypothesize that students' feelings of efficacy and futility are functions of the level of clarity regarding test expectations created by teachers' practices in communicating test expectations and providing feedback about test performance. We also hypothesize that efficacy and futility are functions of the correspondence of tests to those expectations resulting from teachers' practices in constructing tests. Finally, we hypothesize that students' feelings of futility are a function of the degree of teacher helpfulness after students do poorly on tests.

For each hypothesis, we will examine correlations of student-report measures of teaching practices with our indices of Efficacy, Futility, and Effort. Then we will introduce teacher-report measures and examine correlations with student reports and with the criterion variables. The main analyses will employ class means of student responses to questionnaire items, although we will refer to relationships among variables at the individual student level where appropriate.

Student Clarity and Teacher Communication and Feedback

We expected student efficacy and effort to be positively related, and futility to be negatively related, to student clarity about learning objectives. The more definite the student's knowledge of what to study, the more effective should be the study. We expected clarity in turn to be positively related to the student's experience of teacher communication about material to be covered on a test. Furthermore, assuming repeated test experience, we expected clarity to be positively related to the specific--y of feedback of test results regarding material the student still needs to master.

Students' clarity about learning objectives was measured by a single question asking students how often "I know what I am expected to be learning in this class" (item 40). Students' experience of teacher communication about test content was measured by two items asking how often "the teacher gives notice about what will be on a test enough in advance for me to prepare for it" (item 41) and how often "the teacher makes clear the things I should be studying for the test" (item 27). Students' perception of teacher feedback on test results was also measured by two items. They asked the student how often "when I miss something on a test in this class, the teacher gives me specific feedback about what I need to study again" (item 28) and how often "results from the tests in this class let me see easily what I need to review" (item 17). The distributions of student responses on these items are shown in Table III-1.

The data show that the majority of students in each subject often or nearly always knew what they were expected to be learning. Similar majorities reported that the teacher of the class often or nearly always communicated what would be on tests early enough and clearly enough. Test

Table III-1
Students' Reports of Clarity, Teacher Communication, and Teacher Feedback
(Percentage of students in each subject and in all
selecting questionnaire responses)

Questionnaire Item	Response	Biol	Geom	Engl	Hist	All
Clarity:						
40. I know what I'm expected to be learning in this class	Nearly always	38	47	43	42	43
	Often	33	36	32	30	33
	Occasionally	22	13	19	21	18
	Almost never	7	4	6	7	6
Communication:						
41. The teacher in this class gives enough advance notice about tests for me to prepare for them	Nearly always	49	56	56	60	55
	Often	24	22	26	21	23
	Occasionally	19	17	13	13	15
	Almost never	8	6	6	6	6
27. The teacher in this class makes clear the things I should be studying for a test	Nearly always	46	51	50	50	50
	Often	23	31	32	24	28
	Occasionally	21	16	14	18	17
	Almost never	10	3	3	6	5
Communication index (avg. items 41+27)		3.10	3.30	3.31	3.30	3.25
Feedback:						
17. The results from the tests in this class let me see easily what I need to review to get a good grade	Nearly always	28	40	28	28	31
	Often	36	36	34	34	35
	Occasionally	26	18	29	25	24
	Almost never	10	6	9	13	9
28. When I miss things on a test, this teacher gives specific feedback on what to study again	Nearly always	11	20	17	16	16
	Often	25	22	23	20	22
	Occasionally	29	33	32	28	31
	Almost never	35	25	27	35	30
Feedback index (avg. items17+28)		2.47	2.72	2.55	2.48	2.56
Number of students **		390	478	420	405	1694

** Actual number for each item depends on number of missing responses.

results were also seen by most students as informative about what to study to get a good grade. On the other hand, only a minority of students in each subject reported that teachers often or nearly always gave them specific feedback on test items missed. From this last finding, we infer that students are often left to interpret test results by themselves.

As shown in Table III-2, class-level correlations between the two items measuring communication and between the two items measuring feedback were sufficiently strong to warrant averaging each pair of items into indices called Communication and Feedback, which we employ in analyses along with item 40 as an index of Clarity. Means on Communication and Feedback in each subject are shown in Table III-1.

We move now to class-level analyses with means on these indices. First, as the data in Table III-2 show, these three indices were strongly related in each subject. Only the correlation of Clarity to Feedback in English (.31) failed to reach statistical significance. This is evidence that supports our expectations that students' clarity about learning objectives would be positively influenced by the communication and feedback students received about test content. Moreover, students who reported more frequent communication also reported more frequent feedback.

Table III-2 also shows that these indices were strongly related to the Efficacy and Futility indices in biology and geometry and--except for the correlations between Feedback and Futility--in English and history as well. In addition, Clarity and Communication generally were positively related to Effort; Feedback was strongly related to Effort only in geometry. Thus, based on analyses of student questionnaire data, the hypothesis that efficacy and futility are functions of clarity resulting from communication and feedback is confirmed. Furthermore, the positive relationships of Communication and Clarity to Effort suggest that Efficacy and Clarity are

Table III-2
Correlations of Clarity, Communication, and Feedback
with Efficacy, Futility, and Effort
(Class-level correlations in each subject and in all*)

Indices Correlated		Biol	Geom	Engl	Hist	All
Communication items:						
Gives advance notice x Makes clea		.88	.89	.67	.65	.76
Feedback items:						
Gives feedback x Test results inform		.61	.55	.59	.62	.58
Clarity x:	Communication	.68	.[illegible]3	.59	.87	.67
	Feedback	.68	.66	.31~	.57	.58
Communication x Feedback		.84	.80	.52	.68	.69
Efficacy x:	Clarity	.86	.47	.73	.76	.61
	Communication	.79	.70	.55	.66	.65
	Feedback	.75	.79	.51	.44	.53
Futility x:	Clarity	-.49	-.47	-.68	-.45	-.54
	Communication	-.59	-.56	-.66	-.28~	-.49
	Feedback	-.53	-.61	-.10~	.18~	-.27
Effort x:	Clarity	.42	.53	.61	.26~	.40
	Communication	.17~	.68	.55	.39~	.44
	Feedback	.02~	.85	.15~	.28~	.30
Number of classes		16	18	18	17	69

* All correlations except those marked "~" significant at $p<=.05$.

mediating variables in the influence of Communication on Effort.

The strength of these correlations prompted us to investigate whether they might be affected by the segregation of some students into high-track and low-track classes. The Track index was not related to Clarity or Communication and only weakly related (negatively) to Feedback. When we inspected the correlations for "regular" classes alone, we found the same overall pattern of coefficients.

Analyses relying entirely on student reports about teacher practices run the risk of subjective response set, and they also are less informative to teachers than analyses using teachers' own reports of their practices. Hence we turn to teacher questionnaire data on communication and feedback practices to corroborate the findings of the analyses with student data.

Although teachers undoubtedly communicate explicitly and implicitly about test content throughout the conduct of instruction, we focused on specific and deliberate techniques to prepare students for tests. First, we restricted our questions to what we called "unit" tests rather than including quick quizzes or final exams. The teacher questionnaire asked teachers how regularly they communicated test expectations to students at the beginning of, during, and at the end of units (items 21a-c). Table III-3 shows the pattern of teachers' responses in each subject and in the sample as a whole.

The majority of teachers in each subject reported regular use of one or more practices for communicating expectations about tests to students. There were differences among subjects in the percentage of teachers reporting particular practices. For example, 76 percent of the English teachers, compared to only 28 percent of the geometry teachers, reported that they regularly informed students at the beginning of a unit what they would be expected to know on the unit test. This difference may be attributed to the students' unfamiliarity with the content of geometry units, compared to units

Table III-3

Teachers' Communication of Test Expectations and Feedback of Test Results
(Percentage of teachers, in each subject and in all, selecting questionnaire responses)

Questionnaire Item	Response	Biol	Geom	Engl	Hist	All
Communication:						
21. In order to inform students in this class about what they will be expected to know on a unit test, how often do you:						
a. inform students at the beginning of the unit?	Regularly	44	28	76	47	48
	Sometimes	50	39	24	35	37
	Rarely	6	33	0	18	15
b. give students sample test questions during a unit?	Regularly	50	61	52	29	48
	Sometimes	38	33	35	53	40
	Rarely	12	6	12	18	12
c. do a formal review of knowledge, skills just before a unit test?	Regularly	56	83	82	82	76
	Sometimes	38	17	18	12	21
	Rarely	6	0	0	0	3
Feedback						
23c. When scoring unit tests, how often do you provide written comments?	Regularly	19	29	39	35	31
	Sometimes	56	59	44	53	53
	Rarely	25	12	17	12	16
25a. When discussing test results how often you inform the class about items on which many students did poorly?	Regularly	81	89	61	71	75
	Sometimes	19	11	39	29	25
	Rarely	0	0	0	0	0
Number of teachers *		16	18	18	17	69

* Actual number for each item depends on number of missing responses.

in other subjects, until it is explained in class lessons. The finding that more geometry teachers than other teachers reported that they regularly gave students sample test questions during a unit may indicate attempts to compensate for the infrequency of advance communication of test expectations. Geometry teachers (and other teachers as well) built expectations about test coverage during the conduct of a unit. Finally, the large majority of geometry, English, and history teachers reported that they regularly conducted a formal review of unit material at the end of the unit, a day or so before the test. Only in biology did a substantial proportion (44 percent) report that they conducted a formal review only "sometimes" or "rarely."

Teachers not only communicate expectations about tests in advance of each test; they also build expectations about future tests by the kinds of feedback they give regarding student test performance on each test. The questionnaire asked teachers how regularly they provided students with written and oral feedback about test performance. Responses to two questionnaire items are relevant here. First, teachers were asked how regularly, when scoring unit tests, they provided written comments on test papers (item 23c). Second, teachers were asked how regularly, when discussing test results, they informed the class about parts of the test on which many students did poorly (item 25a). As shown in Table III-3, only a minority of teachers in each subject reported regularly providing written comments on tests to supplement or elaborate on test scores, and several teachers reported "rarely" providing written comments. In constrast, the majority of teachers in each subject reported regularly informing their classes about test items on which many students did poorly, although 39 percent of the English teachers reported doing this only "sometimes." No teacher reported providing such oral feedback only "rarely."

In order to see whether these three items on communication and two items on feedback could be combined into composite indices, we looked at the correlations, which are shown Table III-4, among these teacher practices for each subject and for the whole sample of 69 teachers. We found considerable independence among use of these practices. Among the techniques of communicating to students about test expectations, the whole-sample correlation between notifying students at the beginning of the unit and giving sample questions during the unit was only .26. The correlation of these variables reached statistical significance only among English teachers ($r=.73$). The whole-sample correlation between giving sample questions and conducting a formal review at the end of the unit was .33; here, it was in biology alone that the correlation reached significance ($r=.73$). Moreover, there was no significant correlation between beginning-of-unit and end-of-unit notification either at the subject level or the whole-sample level. Likewise, the relationship between the use of written and oral feedback was not statistically significant. Hence we retained the individual items for correlational analyses with student variables.

What was the relationship among these teacher practices and student indices of Communication, Feedback, and Clarity? Table III-4 shows the correlations for the whole sample of 69 classes and for each subject. There were no consistent relationships across subjects. We did find that informing students at the beginning of a unit was positively related to Clarity in geometry, and that giving sample questions during a unit was positively related to Communication in both biology and geometry. In addition, we found that informing the class about test items that many students missed was related to Feedback and Clarity in geometry and history but not in biology and English. Thus we are left with scattered and subject-specific findings regarding the relationship of teacher communication and feedback practices to

Table III-4

Communication and Feedback: Correlations of Teacher Reports and Student Indices
(Class-level correlations in each subject and in all)

Items Correlated			Biol	Geom	Engl	Hist	All
Teacher Reports:							
Communication items:							
	Beginning of unit x:	During unit	.26	.30	.73*	.16	.26*
		End of unit	.17	-.03	.11	.17	.08
	During unit x	End of unit	.73*	-.08	.28	.23	.33*
Feedback items:							
	Written comments x	Discuss items	.20	.40	.09	-.16	.08
Teacher Reports x Mean Student Indices:							
Communication:							
	Beginning of unit x:	Communication	.01	.20	-.07	-.01	.04
		Clarity	-.07	.43*	.24	-.09	.02
		Efficacy	.19	.29	.34	-.18	.21*
		Futility	-.16	-.23	-.26	-.24	-.17
		Effort	.23	.21	-.08	.07	.17
	During unit x:	Communication	.43*	.46*	.11	-.36	.16
		Clarity	.22	.08	.02	-.52*	.00
		Efficacy	.22	.38	.18	-.28	.11
		Futility	.07	-.43*	-.23	.37	-.08
		Effort	.10	.32	-.21	.09	.06
	End of unit x:	Communication	.26	.24	-.20	-.04	.15
		Clarity	.31	.30	-.50*	-.22	.04
		Efficacy	.12	.16	-.13	-.40*	-.05
		Futility	.24	-.07	.15	.41*	.17
		Effort	.49*	.36	-.20	-.07	.18
Feedback:							
	Written comments x:	Feedback	.32	.12	.16	-.34	.06
		Clarity	.29	.09	-.02	-.23	.05
		Efficacy	.52*	.07	.18	-.18	.16
		Futility	-.49*	-.10	-.07	.05	-.14
		Effort	.32	-.08	.06	.14	.12
	Discuss items x:	Feedback	-.24	.56*	.02	.66*	.29*
		Clarity	-.33	.59*	.01	.54*	.17
		Efficacy	-.21	.40*	-.06	.28	.04
		Futility	.20	-.33	.16	.32	.09
		Effort	.25	.39*	.06	.14	.13
Number of classes **			16	18	18	17	69

* $p \leq .05$

** Actual number for each correlation depends on number of missing teacher responses.

students' perceptions of those practices.

We checked to see whether these teacher reports of testing practices related to Efficacy, Futility, and Effort. Table III-4 shows the correlations between the teacher practices and those student indices. The correlations between each of the communication practices and Efficacy in biology and geometry were positive although not statistically significant; the correlation between beginning-of-unit communication and Efficacy for the sample as a whole was statistically significant. The data also show that the correlation between during-the-unit communication and Futility was negative and significant in geometry. These were scattered findings, however. Similarly scattered were findings regarding feedback practices and student indices.

None of the teacher reports was related to the Track index, and the findings in Table III-4 persisted when we restricted our analyses to regular classes alone. In sum, we have failed to identify specific teacher practices that contribute strongly and consistently to student perceptions of good teacher communication and feedback. We have found, however, that such perceptions are positively related to students' feelings of efficacy and effort and negatively related to students' feelings of futility. From this discrepancy in findings, we conclude that further research is warranted on teaching practices that are responsible for helping students to feel clear about what to study.

Before turning to our second hypothesis, we note in Table III-4 that the findings in history are anomalous. In that subject, communication during a unit was significantly <u>negatively</u> related to student Clarity, and communication at end of unit was significantly negatively related to Efficacy and positively related to Futility. This suggests that there may be something dysfunctional about these teacher communication practices in the

U.S. history classes we studied.

<u>Student Perception of Correspondence and Teacher Construction of Tests</u>

Our second hypothesis predicted that efficacy would be a positive function--and futility a negative function--of the degree to which tests in fact cover the material that the student has studied, which we call "correspondence." Correspondence was measured by two items on the student questionnaire: how often tests "cover what I expect them to cover" (item 20) and how often test scores "closely reflect what I have learned" (item 25). Distribution of responses on these items is shown in Table III-5.

In general, the large majority of students in each subject reported that tests often or nearly always covered what they expected and that test results were accurate indicators of learning. The class-level correlations between these two items are shown in Table III-6; they were sufficiently strong to justify averaging the two items into an index called Correspondence.

The relationships between the indices of Correspondence, Efficacy, Futility, and Effort are also shown in Table III-6. They were strong and in the direction hypothesized. They provide evidence that student perceptions of correspondence between what is taught and what is tested have a positive influence on students' beliefs that study pays off and on efforts to study. Correspondence was not related to the Track index, and the strength of relationships was maintained when we restricted analyses to regular classes.

As with communication and feedback, we used the teacher questionnaire to ask about practices that might increase the correspondence of test content to student expectations. We focused on how teachers constructed tests. First, we ascertained whether teachers really "constructed" tests each year. We asked how many of the questions on their tests changed each year (item 31,

Table III-5
Test Construction Practices and Correspondence of Tests and Expectations
(Percentages of teachers and students, in each subject and in all, selecting questionnaire responses)

Questionnaire Item	Response	Biol	Geom	Engl	Hist	All
Student Reports of Correspondence:						
20. The tests given by the teacher in this class cover what I expect them to cover	Nearly always	37	56	50	41	47
	Often	35	28	31	31	31
	Occasionally	20	10	16	22	17
	Almost never	9	5	3	8	6
25. The scores I get on tests in this class closely reflect what I have learned	Nearly always	31	31	35	33	32
	Often	37	38	37	31	36
	Occasionally	22	22	20	26	22
	Almost never	10	10	8	11	10
Correspondence index (avg. items 20+25)		2.95	3.13	3.13	2.94	3.04
Teacher Reports of Test Construction:						
22. When you are constructing a unit test for this class, how often do you:						
c. Use a written list of learning objectives in developing test items?	Regularly	40	44	28	35	37
	Sometimes	40	6	39	41	31
	Rarely	20	50	33	24	32
f. Establish standards that students must meet before they move to a new unit?	Regularly	6	25	11	0	10
	Sometimes	50	31	61	24	42
	Rarely	44	44	28	76	48
Number of students *		391	478	420	405	1694
Number of teachers *		16	18	18	17	69

* Actual number for each item depends on number of missing responses.

Table III-6
Correlations of Test Correspondence and Test Construction Practices
(Class-level correlations in each subject and in all)

Items Correlated		Biol	Geom	Engl	Hist	All
Mean Student Indices:						
Correspondence items: Tests cover what expected x Results reflect what learned		.67*	.81*	.59*	.47*	.57*
Correspondence x:	Efficacy	.79*	.76*	.71*	.62*	.66*
	Futility	-.61*	-.82*	-.70*	-.33	-.62*
	Effort	.43*	.80*	.57*	-.09	.46*
Teacher Reports:						
Use written list x Establish standards		.32	.42*	.22	-.27	.19
Teacher Reports x Mean Student Indices:						
Use written list x:	Correspondence	.36	.15	-.04	.01	.06
	Efficacy	.49*	.25	.17	.15	.22*
	Futility	-.19	-.28	-.08	.30	-.06
	Effort	.53*	.19	-.05	.19	.16
Establish standards x:	Correspondence	.40	.32	.10	.17	.32*
	Efficacy	.49*	.12	-.03	.10	.15
	Futility	-.19	-.07	-.24	-.07	-.20*
	Effort	.18	.51*	.29	.27	.33*
Number of teachers** and classes		16	18	18	17	69

* $p <= .05$
** Actual number for each correlation depends on number of missing teacher responses.

results not tabulated). Many teachers in each subject responded that they changed "most" or "all" of their test questions from year to year; the proportions ranged from two thirds of the English teachers to only one third of the biology teachers. All teachers said they changed at least "some," as opposed to "virtually none," of their test questions each year. Hence our focus on test construction seemed justified.

The questionnaire asked the teachers how regularly, when constructing tests, they used a written list of objectives to write test questions (item 22c). We reasoned that such a list of objectives was likely to help the teacher integrate instruction and assessment. The questionnaire also asked the teachers how regularly, when constructing tests, they established performance standards that students had to meet before moving on to the next unit (item 22f). We reasoned that such standards would lead to correspondence in level of difficulty as well as coverage. Distributions of teacher responses to these two questions are shown in Table III-5.

Roughly similar percentages--from 28 percent to 44 percent--of teachers in each subject reported "regularly" using a written list of learning objectives in developing test items. Greater percentages of geometry teachers than other teachers selected each extreme response. Hence the sample of geometry teachers was split on this measure. Geometry teachers also split more than other teachers on how often they established standards. Far more history teachers than other teachers indicated that they "rarely" established standards.

What was the relationship between these two test construction practices? Table III-6 shows the correlations for the whole sample of 69 teachers and for teachers of each subject. Except in history, the relationship between using a written list of objectives and establishing performance standards was positive, reaching statistical significance in geometry. We chose to retain the separate items rather than combine them into an index.

We next looked at the relationships of these test construction practices to students' perceptions of test correspondence with expectations. Table III-6 shows that establishing performance standards exhibited stronger correlations to Correspondence than using written lists of objectives, and the former correlation was statistically significant in the sample as a whole.

We investigated direct effects of teacher practices on class means of student Efficacy, Futility, Effort; these results, too, are shown in Table III-6. There was a positive and statistically significant relationship between both using written lists and establishing performance standards and Efficacy in biology, but other relationships were weaker. Also, each test construction measure was generally negatively related to Futility, and the correlation between establishing standards and Futility was significant in the sample as a whole. This is interesting because one might have surmised that standards would contribute to perceived futility. We also observe significant positive correlations between using standards and Effort in the sample as a whole and in geometry, and a significant positive relationship between using lists of objectives and Effort in biology.

In sum, as with the analyses for the first hypothesis, we have found only scattered evidence supporting the claim that these testing practices influence student feelings of efficacy, despite the strong relationships of student perceptions of correspondence between expectations and tests with Efficacy and Effort. The specific practices we measured are perhaps not the best indicators of how teachers go about aligning tests with instruction. Such alignment remains an important subject for research.

Student Futility and Teacher Helpfulness

In our third hypothesis, we restricted our attention to students' feelings of futility and posited that we expect that futility would be alleviated by teachers' follow-through by providing help after students had done poorly on tests. We did not expect there to be a relationship between teacher follow-through and students' feelings of efficacy, which as measured depended on what a teacher did before a test rather than after it.

The student questionnaire asked how often, "When a student gets a low score on a test in this class, the teacher makes sure he or she gets the help needed to do better" (item 18). Distribution of student responses on this item, which is our index of teacher helpfulness, is shown in Table III-7. Only in geometry did a majority of students report that teachers often or nearly always provided help; in biology and history, only 40 percent of the students reported thus. Thus helpfulness, like feedback, is experienced less often than communication and correspondence.

Class-level correlations between perceptions of teacher Helpfulness and Futility were negative in each subject. As Table III-8 shows, although the class-level correlations were not significant in English and history, in general, students who perceived that teachers gave them the help they needed felt less hopeless about doing well.

Like most other student perception indices, Helpfulness was not related to the Track index. When we recomputed the correlations for regular classes, the correlations between Helpfulness and Futility remained statistically significant for the whole sample and for biology. Moreover, although the correlation in regular geometry classes fell below the level of statistical significance, it was strong and statistically significant in the low-track geometry classes. (Geometry alone had enough low-track classes to permit use of correlational analyses in that subsample.)

Table III-7
Teacher Helpfulness to Students Performing Poorly on Tests
(Percentages of students and teachers, in each subject and in all, selecting questionnaire responses)

Questionnaire Item	Response	Biol	Geom	Engl	Hist	All
Student Reports:						
18. When a student gets a low test score, the teacher provides the help needed to do better next time	Nearly always	17	26	19	18	20
	Often	20	26	27	20	24
	Occasionally	36	28	33	35	32
	Almost never	28	20	22	27	24
Teacher Reports:						
27. When many students do poorly on a unit test, how often do you:						
a. reteach the material	Regularly	7	29	39	12	23
	Sometimes	73	53	56	62	61
	Rarely	20	18	6	25	17
b. move on to keep to the course schedule	Regularly	20	29	24	29	26
	Sometimes	60	41	47	47	48
	Rarely	20	29	29	24	26
28. When an individual student or small group of students does poorly on a unit test, how often do you:						
a. give special remedial instruction outside class time	Regularly	19	18	17	31	21
	Sometimes	44	82	44	31	51
	Rarely	38	0	39	38	28
b. create special learning activities in class besides regular class work	Regularly	6	12	6	0	6
	Sometimes	44	25	39	50	39
	Rarely	50	62	56	50	54
c. arrange for peer tutors	Regularly	6	18	0	6	8
	Sometimes	50	47	61	31	48
	Rarely	44	35	39	62	45
Number of students *		391	478	420	405	1694
Number of teachers *		16	18	18	17	69

* Actual numbers for each item depend on numbers of missing responses.

Table III-8
Teacher Helpfulness: Correlations of Student Indices and Teacher Reports
(Class-level correlations in each subject and in all)

Indices Correlated		Biol	Geom	Engl	Hist	All
Mean Student Indices:						
Helpfulness x:	Futility	-.44*	-.44*	-.35	-.10	-.36*
	Effort	.13	.69*	.29	.27	.36*
Teacher Reports:						
Reteach x:	Move on	-.42	-.57*	-.67*	-.40	-.52*
	Remediation	.14	-.08	-.09	.36	.08
	Special Activities	.01	.07	.48*	.42	.23*
	Peer Tutoring	.07	.04	.46*	.51*	.25*
Move on x:	Remediation	.00	-.20	.20	.11	.06
	Special Activities	.34	.42*	-.33	.08	.12
	Peer Tutoring	-.18	.11	-.07	-.47*	-.14
Remediation x:	Special Activities	.52*	-.27	.13	.23	.18
	Peer Tutoring	.27	-.10	-.09	-.19	.00
Special Activities x Peer Tutoring		.06	-.30	.47*	.10	.04
Mean Student Indices with Teacher Reports:						
Helpfulness x:	Reteach	.03	.25	.47*	.32	.31*
	Move on	-.36	-.40	-.17	-.05	-.24*
	Remediation	.14	.01	-.29	.71*	.20*
	Special Activities	.02	-.09	.55*	.04	.12
	Peer Tutoring	.64*	.21	.57*	.09	.38*
Futility x:	Reteach	.38	-.02	-.58*	.15	-.07
	Move On	.07	.29	.28	-.31	.07
	Remediation	-.05	.34	.28	-.09	.06
	Special Activities	.20	.09	-.25	.02	.02
	Peer Tutoring	-.41	-.05	-.36	-.20	-.27*
Effort x:	Reteach	.10	.45*	.45*	.02	.33*
	Move On	.16	-.57*	-.11	-.27	-.24*
	Remediation	.73*	-.30	.14	.10	.17
	Special Activities	.36	-.14	.06	-.16	.02
	Peer Tutoring	.51*	.13	.10	.16	.19
Number of teachers* and classes		16	18	18	17	69

* p<=.05
** Actual number for each correlation depends on number of missing teacher responses.

We now turn to teacher questionnaire items regarding a number of practices that provide help to students who do poorly on tests. Although nearly all teachers agreed with a questionnaire item stating that school has a special responsibility to students who are failing (item 6b, results not tabulated), we recognize that the work load of high school teachers and the demands of the curricular schedule place limits on what a teacher can do. We focused on possible responses in two situations: where many students in a class do poorly on a test and where only a few students do poorly.

Teacher follow-through on test failure was measured by two items asking what teachers did when many students did poorly on tests. Distribution of responses is shown in Table III-7. First, teachers might reteach the material missed by many students (item 27a). Only 7 percent of the biology teachers (one teacher) and 12 percent of the history teachers (two teachers), compared with 29 percent of the geometry teachers and 39 percent of the English teachers, retaught the material regularly. Second, teachers might move on to the next unit to keep to the schedule (item 27b). There was no difference across subjects in the percentage of teachers reporting that they regularly moved on to the next unit in order to keep to the course schedule. Moreover, the majority of teachers clearly responded "sometimes" to both options.

The correlation between the two measures of response to widespread student test failure is shown in Table III-8. It is evident that for the sample as a whole, there is a negative correlation (-.52) between tendency to move on and tendency to reteach. This correlation is negative in each subject area, although it reaches statistical significance only among geometry teachers and English teachers.

In addition to responses to classwide test failures, teachers were asked about a number of possible techniques for dealing with <u>individual</u>

students or small groups of students who did poorly on tests (and who didn't require that the teacher delay the course schedule for the class as a whole). These techniques included giving special remedial instruction outside class time, creating special learning activities in class in addition to regular class work, and arranging for peer tutors (items 28a-c). Distribution of teacher responses on these items is shown in Table III-7. What is immediately notable is that never does a majority of the respondents report regular use of any of these practices. In general, teachers are most likely to give special remedial instruction outside class time.

How related are these practices for dealing with small groups of students performing poorly on tests? Table III-8 shows that providing remedial instruction and providing special learning activities are positively related among biology teachers (r=.52), and the latter is positively related to arranging for peer tutoring among English teachers (r=.47). Otherwise, and in general, these practices did not constitute a response syndrome.

Table III-8 also shows that there are positive relationships of reteaching to providing special activities and arranging peer tutoring in English and history. Against this, it must be noted that for geometry teachers there is a relationship between moving on and providing special activities (r=.42). Moving on does show negative relationships to arranging for peer tutoring, except in geometry. As before, it is wiser to consider the practices separately rather than combine them into an index.

How do these teacher-reported practices relate to student reports of teacher helpfulness? The correlations in Table III-8 show that teachers' reports of reteaching after classwide failure are positively related, and moving on negatively related, to students' reports of teacher provision of help. Teacher provision of remedial instruction (except in English) and arranging for peer tutoring are likewise related to student perceptions of helpfulness. Thus, in the area of teacher response to students' poor test performance, where both teachers and students report lower levels of teacher effort than in the case of communication, feedback, and correspondence, there are significant relationships between teacher reports of practice and student reports of perceptions. Evidently students' perceptions are sensitive to teacher efforts. In English, reteaching, providing special learning activities, and arranging for peer tutoring seem to increase student awareness of teacher helpfulness; providing remedial instruction outside class time seems most effective in history, and arranging for peer tutoring seems to increase student awareness most in biology.

But are these teacher practices effective in reducing futility and increasing effort? Table III-8 shows the correlations of Futility and Effort with the five teaching practices. Reteaching is significantly negatively related to Futility only in English. This isolated finding regarding Futility gains significance when we observe that, also in English, reteaching is positively related, and moving on negatively related, to Effort. Table III-8 also shows that reteaching is positively related and moving on negatively related to Effort in geometry and in the sample as a whole. These findings suggest that reteaching when many students do poorly on a test is an effective practice.

With respect to teacher responses to individual test failures, Table III-8 shows that the correlation of peer tutoring to Futility is negative in all cases and statistically significant for the sample as a whole. The strongest subject correlation is in biology (-.41), and we also observe that peer tutoring--along with providing remediation--is significantly positively related to Effort in biology. Therefore, we suggest that peer tutoring may be an effective practice in that subject for reducing feelings of futility and increasing study efforts of individual students who are having difficulty mastering the material.

Summary

We have found abundant evidence that students' perceptions about communication, feedback, correspondence, and helpfulness are strongly interrelated and strongly related to student feelings of the efficacy versus futility of study and to student reports of their own effort in study. On the basis of these findings, we would argue that increasing student perceptions of what we regard as desirable class testing practices is a promising avenue for increasing feelings of efficacy and levels of effort.

Increasing such perceptions depends, however, on increasing certain teacher practices, and we cannot say that the practices we have identified and analyzed are powerful and consistent levers on student perceptions. Moreover, these practices do not seem to differentiate teachers consistently; few of the practices were related to one another nearly as strongly as were class means of student perceptions of testing practices. On the other hand, the relationships we did find should be taken seriously by teachers who may want to try some particular practice that is positively related to student efficacy and/or effort in their subject area.

We expect that there are other, perhaps more diffuse, differences in teaching practices that affect how students perceive and react to tests. It is likely that the informal culture of expectations built up over the year by teacher remarks and reactions operates somewhat independently of the specific practices we have studied. It is possible that the strong relationships among class means of student data are attributable to diffuse differences in teaching practices that heighten student awareness of test expectations and the relationship between test coverage and instructional objectives. We did not ask explicitly about liking for or satisfaction with teachers, but the constellation of strongly-related class means on positively-toned questionnaire items suggests such a "halo" effect. (Such a halo, of course, may be well deserved by the teacher who manages to create such generalized positive perceptions of class tests.)

We are not willing to abandon the specific practices described in this section, however. They remain strongly rooted in instructional theory and rational in terms of the model of student motivation presented here. It may be that few of those practices are really implemented by teachers who responded that they "regularly" used them; teachers' benchmarks of use and of regularity may be highly situation-bound and, in the case of the more ambitious practices such as providing special learning activities, may exaggerate what are minimal efforts compared to the requirements of a mastery learning program. In fact, some measures that had been included in the questionnaire, such as teachers providing subscores on tests to identify mastery of different sets of learning objectives and teachers adapting their instructional goals and tests to the needs of low-achieving students, were dropped from analysis after preliminary inspection. Either too few teachers reported ever using them, or teachers' written comments on the questionnaire revealed interpretations disparate from the significance of such practices in our view of how to integrate teaching and testing. Even if one should be cautious in taking teacher self-reports on these practices at face value, the practices retain an integrity in terms of improving high school instruction, and this warrants continued efforts to identify teachers who are using them and to study their effects on students.

IV. Influence of Policy and Collegiality on Class Testing Practices

The model of the relationship of high school teachers' class testing practices to students' feelings of efficacy and efforts to study was developed in the context of an experimental study of a staff development intervention to increase teachers' use of the practices described in Section III. For a report of the findings of that study see Fielding, Shaughnessy, and Duckworth (1986). Because we were aware that the teaching practices of communication and feedback, test construction of tests around objectives, and response to poor test performance all impose burdens on high school teachers already occupied by the demands of teaching five classes each day, we were interested in the potential of collegial assistance to support increased use of such practices. Hence the intervention included attempts to foster collegial cooperation among teachers in improving the design and conduct of class testing. It also included attempts to generate administrative support for such improvements in teaching practice. The finding of that study--that trained teachers exhibited little change in practice--was attributed in part to the failure of collegial cooperation and administrative support to mature. We concluded that future efforts to improve class testing practices must be rooted in school and departmental policy decisions and must build stronger collegial support than was achieved in the experimental study.

In this section, we present evidence from our comparative data on four school subjects regarding the influence of school and departmental policies and collegial climate on the class testing practices of communication, feedback, purposive construction, and response to student failure. Because of the limited evidence presented in Section III that the specific testing practices measured by our questionnaire affect students' perceptions of the class and students' feelings of efficacy and levels of effort, we also consider direct effects of policy and climate on student variables.

In order to explore such possible influences on teaching practices, we supplemented data from the teacher questionnaire employed in analyses in Section III with interviews with high school administrators and subject-area department chairpersons. Those interviews focused on policies and practices that might affect class testing and on resources that might enable teachers to respond to poor student performance.

We now summarize what we learned from those interviews regarding the differences among the four schools and subjects regarding school and departmental policies. Then we return to the questionnaire data for further analyses.

<u>School Administrative and Departmental Contexts of Class Testing</u>

In discussions with administrators, it became very clear that class testing was simply not a concern of administrators. The only time they would become involved with tests was in disputes over grading, which were rare. None of the schools or departments had a policy about the format, coverage, or scoring of tests. Tests were within the domain of teachers' autonomy. Furthermore, testing was not generally perceived as a problem area. Evidence of poor student performance on tests tended to be treated less as a stimulus to instructional problem-solving than as an occasion to make moral judgments about the consequences of student laziness and lack of study and to terminate a student's enrollment in a particular course of study and place him or her in a less demanding course.

As a result of District B's policies on teacher evaluation and staff development, the interaction of administrators and teachers around instructional concerns may have been developed more completely in the two schools in that district than in the two schools in District A. Several people in District B talked about its three-stage teacher evaluation cycle, in which each teacher was evaluated cyclically on common teaching criteria, specific criteria selected by the evaluator, and specific criteria selected by the teacher. This process of interaction was intensified in School 3 by the formal involvement of department chairs in evaluation, which was unique in the four schools under study. In contrast, there was little discussion of teacher evaluation by interviewees in District A. In any case, the formal evaluation procedure seems not to have focused on testing practices.

In addition to its systematic evaluation cycle, District B had fostered the implementation of a comprehensive staff development program in teaching skills, whereas District A was only in the beginning stages of such a program. Hence administrators and teachers had been interacting for several years about instructional concerns in District B schools, especially School 4, whereas such interaction was more limited in the two District A schools. However, the staff development programs in question did not focus specifically on testing practices, although they did focus on some of the preparatory teaching behaviors, such as communication of objectives.

In sum, the school administrative context of course testing was benign and nondirective. There was little evidence that schools had attempted to influence this aspect of teaching by administrative means.

In interviews with chairpersons of the science, math, language arts, and social studies departments in each school, we inquired about departmental policies and collegial climate regarding class testing. The general picture was that there were no such policies. This picture was varied only in the case of specific courses that had recently undergone redevelopment, and even then the department seemed only to provide options and guidelines for teachers rather than specify particular practices. Department chairs were often emphatic about the prerogative of the individual teacher regarding judgments of the quality of student performance in scoring tests. In general, each teacher was free to set his or her own testing policy. There were cases of coordination among teachers of multiple-section courses, but department-level policy was nonexistent. The sentiments of many chairpersons were summed up in one science chair's reaction to our probing about departmental meeting discussions of testing issues: "I wish we had the time to go into those issues, because they're important. But we only meet once or twice a month and there are always budget, logistical, and state standardization matters to resolve."

With respect to departmental collegial climate, there were differences among the subject areas. The science and social studies departments were structured as loose federations. In science, there was a clearcut separation of the "life sciences"--predominantly biology--from the "physical sciences" in terms of teachers' certification and subject-matter identity; as a consequence, the department was unlikely to function as a unit with regard to particular courses. In none of the science departments was the chair a biology teacher. Likewise, in social studies departments, teachers of history, government, and economics courses were likely to identify with different academic disciplines and recall different preservice and inservice training experiences. Two of the department chairs were history teachers and talked easily about that subject; two were government teachers and talked very tentatively about history. This intradepartmental division was not apparent in math or language arts, however. Nonetheless, the chairs in those departments acknowledged that the governance of specific courses was likely to devolve on the teachers of those courses.

The math chairpersons went into more detail about curriculum and instruction than did the science chairpersons, and it was evident that each math chair felt responsible for and comfortable with each course in the department. Math is not a federation of different disciplines like science. Moreover, perhaps because the skills taught in the sophomore year are still critical in the senior year, there was more emphasis on mastery at each level. The School 2 chair was most emphatic about this. It may have been relevant that he had recently come to the high school from many years of teaching at the middle-school level.

There were course outlines on file for each math course, but these were regarded as relics of recurrent periods of curricular codification (the most recent having been the articulation of high school graduation competencies) rather than living forces in the conduct of courses. The text was the backbone of the course, although teachers were free to omit chapters, reorder chapters, and supplement the text as they saw fit.

Some of the language arts departments had established detailed outlines for each unit in sophomore English; School 4's department chair emphasized this effort. School 1 used the same final exam in all three le of sophomore English, but unit testing was up to the individual teacher.

When asked about resources for teachers with failing students, most department chairs responded with blank looks. There was the standard opportunity of the student to seek out the teacher during the lunch break or after school. There was also the possibility of altering the student's placement. Beyond this, the school offered little assistance. An exception to this generalization was School 1, where each department had a special resource center, always staffed by a teacher, which students could visit during one of their free periods each day.

Thus we came to regard the department as such as a relatively weak influence on teachers' class testing practices. Our focus came to rest on a different basis for policy and collegiality--the faculty teaching a particular course. We heard several anecdotes about team or cooperative work relationships among such teachers. For example, according to some science chairpersons, biology teachers would voluntarily share course outlines and materials, sometimes across all sections, sometimes in pairs. Hence we turn to the question of variation among teachers in interaction with colleagues and to its relationship to variation in teachers' class testing practices and the consequent student perceptions investigated in Sections II and III.

Collegiality among Teachers of the Same Course

Given the apparent fruitlessness of investigation of departmental influences on testing practices, we follow the lead of several chairpersons and attempt to build a picture of collegial influences on the foundation of individual teacher affiliations, especially within the faculty teaching each course.

The teacher questionnaire asked several items about the working relationship among teachers of the same course. First, teachers were asked whether other teachers also taught the course about which they were answering the questionnaire. If so, they were asked how often they talked to other teachers of that course about course goals, course tests, and grading standards. The distribution of teacher responses to these items in each subject area is shown in Table IV-1. These three items were strongly intercorrelated, as shown in Table IV-2, and were averaged to create an index called Interaction. As the subject means on this index show in Table IV-1, biology teachers had the highest Interaction score and history teachers had the lowest.

In addition to these general measures of collegiality among teachers of the same course, we asked about consultation with other teachers over

Table IV-1
Teachers' Reports of Collegiality
(Percentage of teachers, in each subject and in all, selecting questionnaire responses)

Questionnaire Item	Response	Biol	Geom	Engl	Hist	All
10. How often do you talk to other teachers of this course about your:						
a. Course goals?	Very Often	20	6	0	6	8
	Often	53	31	40	6	32
	Occasionally	20	44	60	71	49
	Seldom/never	7	19	0	18	11
b. Course tests?	Very Often	27	0	0	12	10
	Often	20	38	13	0	18
	Occasionally	40	50	73	53	54
	Seldom/never	13	13	13	35	19
c. Grading standards?	Very Often	21	6	0	12	12
	Often	29	19	13	12	18
	Occasionally	43	50	73	53	55
	Seldom/never	7	25	13	24	18
Interaction index (avg. items 10a-c)		2.71	2.19	2.13	2.00	2.24
22a. When constructing a test, how often do you seek advice from teachers about items?	Regularly	12	6	11	6	9
	Sometimes	44	78	50	41	54
	Rarely	44	17	39	53	38
27c. When many do poorly on a test, how often do you consult other teachers?	Regularly	33	19	6	12	17
	Sometimes	40	69	67	53	58
	Rarely	27	12	28	35	26
28f. When a few do poorly on a test, how often do you talk to other teachers?	Regularly	25	12	19	19	18
	Sometimes	62	82	69	62	69
	Rarely	12	6	12	19	12
9. Do you and other teachers of the course in question:						
a. Use the same text?	Yes	93	93	93	65	86
b. Use the same course outlines?	Yes	73	71	60	44	62
c. Use the same tests?	Yes	46	44	15	19	30
Commonality index (wtd. sum items 9a-c)		3.85	3.86	2.46	2.06	3.02
Number of teachers *		16	18	18	17	69

* Actual number for each item depends on number of missing responses.

Table IV-2
Correlations among Teachers' Reports of Collegiality
(Correlations in each of four subjects and in all)

Items Correlated	Biol	Geom	Engl	Hist	All
Interaction items:					
Talk about goals x: Talk about tests	.83*	.68*	.53*	.67*	.72*
Talk about standards	.89*	.62*	.53*	.76*	.73*
Talk about tests x Talk about standards	.86*	.77*	.50*	.89*	.81*
Interaction x: Seek advice on questions	.68*	.39	.64*	.80*	.60*
Consult on many failures	.61*	-.19	.30	.67*	.44*
Talk about indiv. probs.	.54*	-.06	.73*	.52*	.41*
Seek advice x: Consult on many failures	.54*	.02	.14	.62*	.39*
Talk about indiv. probs.	.55*	-.32	.26	.50*	.33*
Consult x Talk about indiv. probs.	.45*	.25	.68*	.77*	.56*
Commonality items:					
Use same text x: Use same outline	.45*	.44	-.22	.42*	.34*
Use same tests	.27	.25	.12	.37	.29*
Use same outline x Use same tests	.51*	.32	-.03	.54*	.40*
Commonality x: Interaction	.61*	.43	.38	.62*	.58*
Seek advice on questions	.31	.45*	.51*	.54*	.48*
Consult on many failures	.30	-.10	.29	.15	.23*
Talk about indiv. probs.	.12	-.06	.17	.15	.09
Number of teachers **	16	18	18	17	69

* p<=.05
** Actual number for each correlation depends on number of missing responses.

specific testing issues. The questionnaire asked about three occasions for consultation. In connection with reporting their practices in test construction, teachers were asked how regularly they sought advice from other teachers about test items (item 22a). They were also asked how often they consulted other teachers when many students did poorly on a class test (item 27c). Finally, teachers were asked how regularly they talked with other teachers about individual students' learning problems when those students did poorly on tests (item 28f). Teachers were asked to answer these three questions regardless of whether they alone taught the course in question. (In fact, the teachers who said they taught the course alone were all responding in reference to a high-track or low-track class; in each case, there were other teachers of the subject in their department who could provide collegial assistance.)

Table IV-1 also shows distribution of responses on items measuring specific consultation practices. Very few teachers in any subject reported seeking advice in test construction as a regular practice. The subject with the highest percentage of teachers who reported seeking advice "regularly" or at least "sometimes" was geometry; the subject with the lowest percentage of teachers checking those two responses was history, more than half of whom reported seeking advice "rarely." Slightly greater percentages of teachers in each subject reported consulting another teacher when many students did poorly on a test. Here, geometry teachers again were most likely, and history teachers least likely, to consult another teacher at least sometimes, although biology teachers were most likely to consult regularly. This pattern was repeated in teachers' responses to how often they talked to other teachers about learning problems of individual students who had done poorly on a test, although teachers were less likely to say they did this rarely than they were with consultation over classwide testing failures. These items were moderately correlated but are retained as separate items for analysis with measures of the appropriate specific testing practices. The relationship of each of these items to the Interaction index is shown in Table IV-2. All three were consistently related positively to the index, except that the correlation of Interaction to the two items referring to consultation after test failures was weakly negative among geometry teachers.

Pursuant to our interest in policy as a basis for improved teaching practice, we were curious whether course colleagues developed common curricula and instructional tools. Perhaps such decisions about instructional policy might increase teacher interdependence and shared experience, leading to sustained collegial interaction. We had no questionnaire data about the development of course materials, but we did have teachers' responses to questions about whether they and other teachers of the course under consideration used the same text, the same course outlines, and the same tests (items 9a-c). Percentages of teachers responding "yes" in each case are shown in Table IV-1.

It is apparent that use of a common text was prevalent in all courses except U.S. history. The majority of teachers of all courses except history also reported using the same course outlines. In no subject did a majority of teachers report using the same tests, however. In fact, this practice was rare in sophomore English and U.S. history. Correlations among these yes/no items -- shown in Table IV-2 -- were universally positive, but the relationships between using the same text and using the same tests were not statistically significant. Given that frequency of collegial interaction should be most influenced by using the same tests, and least strongly by using the same text (a decision typically made only every five years), we constructed from these items a weighted additive index, called Commonality, in which teachers received 1 point for using the same text, 2 points for using the same course outlines, and 3 points for using the same tests. Mean scores on Commonality, shown in Table IV-1, revealed that biology and geometry teachers were most likely to use common materials, while history teachers were least likely.

Table IV-2 shows the correlation of the Commonality index with each measure of frequency of collegial interaction. Commonality is significantly related to the Interaction index and to the frequency of discussing test items with other teachers while constructing tests. When we probed correlations among individual items, it was apparent that using the same test indeed had the greatest influence on interaction. Table IV-2 also shows that Commonality had a weak relationship (significant only at the whole-sample level) to the frequency of consulting other teachers when many students did poorly. Furthermore, there was no appreciable relationship to discussing individual students' poor test performance with other teachers. In sum, there is evidence that collegial interaction in general is more frequent where teachers have agreed to use common curricula and tests, but collegial interaction over cases of student test failure seems to derive from other sources.

The main question for these analyses was whether collegiality was related to the use of the testing practices described in Section III. Table IV-3 shows the correlations of the Interaction index and, where appropriate, specific items to measures of those testing practices. With respect to measures of communication and feedback, there were no significant correlations with the Interaction index at the whole-sample level, although in biology there were positive relationships between Interaction and all five measures, reaching statistical significance in the case of holding end-of-unit reviews and providing written feedback. A similar pattern of

Table IV-3
Correlations of Collegiality and Teaching Practices
(Correlations in each of four subjects and in all)

Items Correlated	Biol	Geom	Engl	Hist	All
Interaction x:					
Communication:					
beginning of unit	.34	-.15	.04	.34	.10
during unit	.44	-.15	.35	-.17	.09
end of unit	.66*	-.26	.29	.12	.04
Feedback:					
written comments	.60*	-.26	.53*	-.42*	.01
discuss items missed	.16	-.45*	.37	.28	.17
Test construction:					
use objectives	.65*	.27	.40	.15	.41*
establish standard	.43	.08	.51*	-.36	.14
Response to many failures					
reteach	-.12	-.45*	.39	.04	-.12
move on	-.31	.46*	-.28	-.25	-.08
Response to few failures:					
remediation	.16	.14	.23	-.13	.03
special activities	.04	-.12	.30	-.26	-.04
peer tutoring	.43	.09	.34	.30	.29*
Consultation x Test Construction:					
Seek advice on questions x:					
use list of objectives	.32	.49*	.30	.38	.32*
establish standards	.32	-.03	.45*	-.48*	.18
Consultation x Response to Failures:					
Consult on many failures x:					
reteach	.02	.46*	.60*	.24	.28*
move on	-.27	-.43*	-.18	-.10	-.23*
Talk about indiv. probs. x:					
remediation	.34	-.07	.19	.37	.25*
special activities	.15	.31	.45*	.00	.22*
arrange peer tutoring	.30	.24	.33	.17	.25*
Number of teachers **	16	18	18	17	69

* p<=.05
** Actual number for each correlation depends on number of missing responses.

positive relationships was observed in English, where only the relationship with providing written feedback was significant. Relationships were mixed in history, where the only significant relationship was negative, indicating that more frequent interaction was associated with less frequent written feedback. This anomaly was more pronounced in geometry, where all the correlations were negative, the correlation with providing oral feedback significantly so. This suggests that while collegiality in biology may reinforce the communication and feedback strategies studied here, collegiality in geometry may undermine them.

We turn next to teachers' practices in constructing tests. Teachers' reports of using written lists of objectives in constructing tests were significantly related to the Interaction index (r=.41) and the item on seeking advice on test questions when constructing tests (r=.32). Since such lists are teaching tools like the course outlines included in the index of Commonality, this finding suggests that lists are examples of emergent curricular policy among interacting course colleagues. Teachers' reports of establishing performance standards for tests, however, were not consistently related to either measure of collegiality. Given the findings in Section III that performance standards were more strongly related to student perceptions than lists of objectives, this is disappointing. Perhaps performance standards are not usually codified and shared with other teachers, a speculation consistent with department chairs' remarks that judgments about student performance were even more closely guarded as professional prerogatives than the design of tests.

The third area of testing practice under study was teachers' response to poor student performance on tests, which we differentiated into widespread and isolated student test failure. In the case of widespread failure, teachers' reports of reteaching material were positively related, and teachers' reports of moving on to adhere to the course schedule negatively related, to the frequency of consulting other teachers in such cases. The direction of correlation was consistent across subjects, although only in geometry as well as the whole sample were the coefficients statistically significant. Given this finding in geometry, it was very surprising to see that each of these teaching practices was significantly related the opposite way to the Interaction index, where more frequent interaction was associated with moving on rather than reteaching. It is recalled that consultation of other teachers was not related to Interaction in geometry. It is also recalled that Interaction was negatively related to all communication and feedback measures in geometry. This is further evidence that general collegiality in that subject may undermine the teaching practices herein advocated, even though specific consultation in response to widespread test failure sustains the practice of reteaching.

In the cases where individual students did poorly on tests, all three measures of teacher response--providing remedial instruction outside class, providing special learning activities in class, and arranging for peer tutoring--were positively related at the whole-sample level to the frequency of discussing such students' learning problems with other teachers. The correlations, although statistically significant, were weak, however, and in general the subject-level correlations were not statistically significant. Nor were there consistent relationships to the Interaction index, which was related significantly only to teachers' reports of arranging peer tutoring. As in the case of widespread test failure, specific collegial interaction seems more potent an influence on teacher response than does general collegiality.

Given the findings in Section III that student perceptions of teacher communication, feedback, correspondence, and helpfulness were far stronger correlates of students' feelings of efficacy and effort than were the teachers' reports of such practices, we speculated that the specific teaching practices included in the teacher questionnaire may not encompass the ways in which a collegial climate influences teachers' integration of teaching and testing for students. Hence we probed relationships of collegiality to the student variables included in the model. Table IV-4 shows the correlations of the Interaction index and the three specific measures of collegiality to those student variables.

The Interaction index exhibited statistically significant relationships to student perceptions only in biology and geometry. However, the relationships were opposite from one another. In biology, there were positive relationships between Interaction and student perceptions of Communication, Feedback, Clarity, and Correspondence and student feelings of Efficacy. This accords with some science department chairpersons' emphasis on curriculum-building among biology teachers as a positive factor in instruction. In contrast, in geometry, there were negative (and statistically significant) relationships between Interaction and students' perceptions of Feedback and Correspondence and students' feelings of Efficacy. There was also in geometry a positive (and statistically significant) relationship between Interaction and students' feelings of Futility. This accords with the picture we have developed in this section of the adverse effect of collegiality on the integration of teaching and testing in geometry. It is possible, of course, that hidden factors are producing the surprising correlations in geometry; we ruled out one, however, in our findings that tracking, most pronounced in geometry in this study, was not the source of these findings.

With respect to the more specific teacher consultation practices, we found that the practice of seeking advice on test questions was positively

Table IV-4
Correlations of Teacher Collegiality Reports and Mean Student Indices
(Correlations in each of four subjects and in all)

Items Correlated	Biol	Geom	Engl	Hist	All
Interaction index x:					
Communication	.61*	-.38	.20	.15	.13
Feedback	.54*	-.55*	.02	.32	.10
Clarity	.64*	-.20	-.19	.25	.14
Correspondence	.69*	-.54*	.16	-.12	.02
Helpfulness	.41	.00	.20	.35	.19
Efficacy	.64*	-.55*	-.11	.15	.08
Futility	-.24	.45*	.15	.17	.09
Effort	.40	-.38	.02	.23	.02
Seek advice on test items x:					
Communication	.45*	.08	.14	-.23	.14
Feedback	.18	.09	-.09	.11	.13
Clarity	.57*	.41*	-.01	-.16	.22*
Correspondence	.56*	.00	.21	-.37	.16
Helpfulness	.21	.18	.05	.17	.19
Efficacy	.41	.02	-.04	-.23	.04
Futility	.02	-.16	-.26	.28	-.06
Effort	.35	.15	.30	.02	.21*
Consult on many failures x:					
Communication	.54*	.21	.55*	.12	.32*
Feedback	.32	.50*	.39	.55*	.44*
Clarity	.30	.40	.19	.23	.26*
Correspondence	.56*	.39	.43*	.05	.32*
Helpfulness	.24	.47*	.40*	.41*	.25*
Efficacy	.37	.18	.43*	.30	.27*
Futility	-.21	-.07	-.42*	.20	-.12
Effort	.24	.61*	.41*	.21	.30*
Talk about indiv. probs. x:					
Communication	.22	-.07	.11	.28	.14
Feedback	.11	.06	.12	.54*	.23*
Clarity	.63*	-.25	-.04	.28	.20
Correspondence	.32	-.02	.20	-.04	.12
Helpfulness	.23	.05	.30	.46*	.26*
Efficacy	.33	.11	.25	.34	.25*
Futility	-.11	.13	-.05	.17	.03
Effort	.52*	.01	.05	.46*	.25*
Number of teachers ** and classes	16	18	18	17	69

* p<=.05
** Actual number for each correlation depends on number of missing teacher responses.

related at the whole-sample level to the frequency of discussing such students' learning problems with other teachers. The correlations, although statistically significant, were weak, however, and in general the subject-level correlations were not statistically significant. Nor were there consistent relationships to the Interaction index, which was related significantly only to teachers' reports of arranging peer tutoring. As in the case of widespread test failure, specific collegial interaction seems more potent an influence on teacher response than does general collegiality.

Given the findings in Section III that student perceptions of teacher communication, feedback, correspondence, and helpfulness were far stronger correlates of students' feelings of efficacy and effort than were the teachers' reports of such practices, we speculated that the specific teaching practices included in the teacher questionnaire may not encompass the ways in which a collegial climate influences teachers' integration of teaching and testing for students. Hence we probed relationships of collegiality to the student variables included in the model. Table IV-4 shows the correlations of the Interaction index and the three specific measures of collegiality to those student variables.

The Interaction index exhibited statistically significant relationships to student perceptions only in biology and geometry. However, the relationships were opposite from one another. In biology, there were positive relationships between Interaction and student perceptions of Communication, Feedback, Clarity, and Correspondence and student feelings of Efficacy. This accords with some science department chairpersons' emphasis on curriculum-building among biology teachers as a positive factor in instruction. In contrast, in geometry, there were negative (and statistically significant) relationships between Interaction and students' perceptions of Feedback and Correspondence and students' feelings of

Efficacy. There was also in geometry a positive (and statistically significant) relationship between Interaction and students' feelings of Futility. This accords with the picture we have developed in this section of the adverse effect of collegiality on the integration of teaching and testing in geometry. It is possible, of course, that hidden factors are producing the surprising correlations in geometry; we ruled out one, however, in our findings that tracking, most pronounced in geometry in this study, was not the source of these findings.

With respect to the more specific teacher consultation practices, we found that the practice of seeking advice on test questions was positively related at the whole-sample level to students' perceptions of Clarity about what they were expected to learn and to students' reports of Effort. As with the findings with the Interaction index, these two correlations were strongest in biology. However, unlike the Interaction findings, no contrary correlations were observed in geometry, where the sign of each of the (very low) correlations was the same as in biology and where the correlation with Clarity was statistically significant.

Particularly impressive were the findings regarding teachers' reports of consulting other teachers about classwide test failures. This measure was significantly and positively related at the whole-sample level to students' perceptions of communication, feedback, correspondence, helpfulness, efficacy, and effort. Furthermore, these correlations were fairly consistent and often statistically significant in each of the subjects. That these correlations are stronger than correlations of consultation with teachers' reports of specific practices and are not restricted to students' perceptions that teachers gave them the help they needed to do better after performing poorly on a test suggests that this teacher practice signifies some approach to instruction that pays off handsomely in terms of student response.

Finally, teachers' reports of talking to other teachers about individual students' learning problems also exhibit positive correlations to student perceptions of feedback, helpfulness, efficacy, and effort. In this case, however, the whole-sample correlations are weaker than in the preceding case and the subject-level correlations are rarely significant.

Overview of Findings and Implications for Research and Practice

In this final section, we have extended the model of influences on students' feelings of efficacy and level of effort to include the working relationships among teachers. We found no indications of instructional policy at either the school or departmental level that would promote the class testing practices that were the focus of this study. Moreover, we found collegial interaction at the department level to be limited, often limited to administrative matters or compliance with state requirements, and handicapped in science and social studies by the diverse disciplinary loyalties of teachers. Therefore, we focused our investigation on the working relationships of teachers, particularly among teachers who taught the same course.

Collegiality among teachers was measured in terms of frequency of talk among teachers of the same course (the Interaction index) and frequency of specific consultation practices. We found Interaction and consultation on test construction related to Commonality in use of text, course outline, and tests, and we speculated that the testing policy that emerges at the level of teachers of the same course may sustain collegial interaction. However, relationships among Interaction and consultation measures and class testing practices proved inconsistent, although some of the specific consultation measures seemed to have a more generally positive relationship to testing practices than did the Interaction index. As we explored the relationship of collegiality to the student variables, however, we found more consistent effects, except in geometry. In fact, one of the consultation measures--consultation in response to classwide test failures--exhibited stronger overall relationships to student variables than did teacher reports of testing practices discussed in Section III. We speculate that this measure is tapping some characteristic of teachers--perhaps conscientiousness, problem-solving orientation, or openness to learning from mistakes--that increases the effectiveness of class testing without necessarily affecting the specific testing practices we measured. That this consultation measure did not specify that the colleague taught the same course suggests that collegiality based on affinity may be at least as productive as collegiality based on teaching the same course.

The strength of the relationships among the student variables presented in Sections II and III warrant further research on teacher practices that influence those variables. The limited evidence that the teaching practices measured in this study are influential needs to be followed up by experimental manipulation of those practices that goes further than the staff development program described in the report by Fielding, Shaughnessy, and Duckworth (1986). Teaching practices that increase students' awareness of feedback and expectation of teacher help are particularly important. Increasing teachers' commitment to providing feedback and help, and increasing students' commitment to using such feedback and help to increase their mastery of material inadequately learned the first time around, seem logical targets for experimentation. Clearly, such commitment will depend on greater resources than are available, given present school arrangements and work loads.

The evidence that teacher collegiality is related to student feelings of efficacy and levels of effort suggests that school and departmental leadership is needed to develop such collegial interaction. This study does not indicate the advisability of attempting to impose school or departmental testing policies, however, because teachers clearly would resist regimentation of what they regard as a professional right. Instead, encouraging and allowing teachers to collaborate with their colleagues to bring more rationality to their testing may be a more fruitful strategy for increasing students' feelings of efficacy about studying for tests and their consequent level of effort.

References

Fielding, Glen, and Schalock, H. Del. Integrating Teaching and Testing: A Handbook for High School Teachers. Monmouth, Oregon: Teaching Research Division of Oregon State System of Higher Education, January, 1985.

Fielding, Glen; Shaughnessy, Joan; and Duckworth, Kenneth. "Effects on Teacher Practice of a Staff Development Program for Integrating Teaching and Testing in High School Courses." Final Report. Eugene, Oregon: Center for Educational Policy and Management, University of Oregon, 1986.

Fisher, Charles; Berliner, David; Filby, N.; Marliave, R.; Cahen, L.; and Dishaw, M. "Teaching Behaviors, Academic Learning Time, and Student Achievement: An Overview." In Time to Learn, edited by Carolyn Denham and Ann Lieberman. Washington, D.C.: National Institute of Education, 1980.

Fitzpatrick, Kathleen, and Charters, W. W., Jr. "A Study of Staff Development Practices and Organizational Conditions Related to Instructional Improvement in Secondary Schools." Final Report. Eugene, Oregon: Center for Educational Policy and Management, University of Oregon, 1986.

Intili, Jo-Ann. "Structural Conditions in the School That Facilitate Teachers' Practice of Reflectiive Decision Making." Paper presented at the annual meeting of the American Educational Research Association, 1977.

Lawler, Edward, III. "Control Systems in Organizations." In Handbook of Industrial and Organizational Psychology, edited by Marvin Dunnette. Chicago: Rand McNally, 1976.

Natriello, Gary. "Evaluation Processes and Student Disengagement from High School." Paper presented at the annual meeting of the American Sociological Assocation, 1983.

Natriello, Gary. "Organizational Evaluation Systems and Student Disengagement in Secondary Schools." Final Report. St Louis, Missouri: Washington University, 1982.

Purkey, Stewart, and Smith, Marshall. "School Reform: The District Policy Implications of the Effective Schools Literature." Elementary School Journal 85 (1985): 353-389.

APPENDIX A

1. Student Questionnaire
2. Teacher Questionnaire

High School Class Testing Study--Student Questionnaire

Dear Student: We are conducting a study of class testing practices in several Oregon high schools. We would like to know what you think about work and tests in the class you are in this period. No one at your school will ever learn how you personally answered this questionnaire, so please answer honestly and completely. We will share summaries of class responses with teachers in the spring to help them think about their testing practices.

Questions are printed on both sides of this page. Mark your answer with a #2 pencil on the separate answer form. Next to the number for each question, please fill in the circle under the letter that matches your answer.

1. What grade are you in?

A. 9th
B. 10th
C. 11th
D. 12th
E. Other

2. Are you male or female?

A. Male
B. Female

3. What are your main classes?

A. College preparatory subjects
B. Other subjects
C. Don't know

4. What do you plan to do after high school?

A. No special plans
B. Get a full-time job or join military
C. Go to a 2-year college or voc. program
D. Go to a 4-year college
E. Other

5. How far did your parents or guardians go in school?

A. Neither finished high school
B. One or both graduated from high school
C. One or both attended college
D. One or both graduated from college
E. Don't know

6. What was your grade point average for all your classes last year?

A. 3.6 - 4.0 (A- to A)
B. 2.6 - 3.5 (B to B+)
C. 1.6 - 2.5 (C to C+)
D. Below 1.6 (D+ or below)
E. Don't know

For questions 7-15, please select the most accurate answer for the class you are in this period.

7. Which statement best describes your reason for taking this class?

A. It is required or was recommended by a counselor or teacher
B. I wanted to take it
C. Don't know

8. What is the lowest grade you would be satisfied with in this class?

A. A
B. B
C. C
D. D
E. Don't know

GO TO QUESTION 9 ON THIS SIDE

9. For this class, how many hours a week do you spend outside class time studying or doing homework?

A. 1 hour at most a week
B. About 2-3 hours a week
C. About 4-5 hours a week
D. More than 5 hours a week
E. Don't know

10. How important to you is it to do well in this class?

A. Not important
B. Somewhat important
C. Very important

BE SURE TO GO TO ITEM 11 ON YOUR ANSWER FORM.

11. How much do you agree or disagree with this statement: "No matter how hard I work in this class, I can't do as well as I would like."

A. Strongly agree
B. Agree
C. Disagree
D. Strongly disagree

12. How useful after high school do you expect what you learn in this subject to be?

A. Not useful
B. Somewhat useful
C. Very useful
D. Don't know

13. How often would you say you have skipped or cut this class?

A. Hardly ever
B. About once a month
C. About once every couple of weeks
D. About once a week or more
E. Don't know

14. How often are you late to this class?

A. Hardly ever
B. About once every couple of weeks
C. About once a week
D. More than once a week
E. Don't know

15. On how many of the tests in this class so far have you done well?

A. Hardly any of the tests
B. Half of the tests at most
C. More than half of the tests
D. Nearly all of the tests
E. Don't know

PLEASE TURN TO THE OTHER SIDE WHEN YOU FINISH THIS SIDE

Please decide how often statements 16-41 are true in the class you are in this period. Mark the matching circle on the answer form. Be sure to start with item 16 on the form. If you don't know the answer to a question, skip it.

	Almost Never	Occasion-ally	Often	Nearly Always
16. I work hard in this class.	A	B	C	D
17. The results from the tests in this class let me see easily what I need to review.	A	B	C	D
18. When a student gets a low score on a test in this class, the teacher makes sure he or she gets the help needed to do better.	A	B	C	D
19. Assignments in this class have to be done in too short a time.	A	B	C	D
20. The tests given by the teacher in this class cover what I expect them to cover.	A	B	C	D
21. I have my homework done when I come to this class.	A	B	C	D
22. When a student gets a low score on a test in this class, the teacher gives makeup work.	A	B	C	D
23. When I'm not understanding the material for this class, I seek the teacher's help after class or outside class time.	A	B	C	D
24. My mind wanders during lessons in this class.	A	B	C	D
25. The scores I get on tests in this class closely reflect what I have learned.	A	B	C	D
26. I understand what the teacher is talking about during this class.	A	B	C	D
27. In this class, the teacher makes clear the things I should be studying for the test.	A	B	C	D
28. When I miss something on a test in this class, the teacher gives me specific feedback on what I need to study again.	A	B	C	D
29. When a student gets a low score on a test in this class, the teacher has him or her work with another student to learn the material.	A	B	C	D
30. I ask questions in lessons in this class when I don't understand.	A	B	C	D

BE SURE TO GO TO ITEM 31 ON YOUR ANSWER FORM.

	Almost Never	Occasion-ally	Often	Nearly Always
31. Before a test in this class, I study the material thoroughly.	A	B	C	D
32. If I study hard for this class, the effort is rewarded.	A	B	C	D
33. It's possible to get the teacher of this class to reduce the amount of material to be covered on a test.	A	B	C	D
34. When a student in this class shows that a test question was unclear or unfair, the teacher revises test scores.	A	B	C	D
35. I can predict how well I will do on a test in this class based on how hard I have studied.	A	B	C	D
36. The work in this class is difficult for me.	A	B	C	D
37. The tests in this class are harder than the teacher's classwork and homework assignments.	A	B	C	D
38. Doing well on tests in this class depends on factors outside my control.	A	B	C	D
39. I find the subject matter interesting in this class.	A	B	C	D
40. I know what I'm expected to be learning in this class.	A	B	C	D
41. In this class, the teacher gives notice about what will be on a test enough in advance for me to prepare for it.	A	B	C	D

THANK YOU. PLEASE RETURN YOUR ANSWER FORM TO THE MANILA ENVELOPE PROVIDED TO THE TEACHER.

CEPM PROJECT ON TESTING IN HIGH SCHOOL COURSES
TEACHER QUESTIONNAIRE

Dear Teacher: These questions about your practices in course testing ask you to select one of several responses to each item or, in a few cases, to provide a piece of information. Mark your answers on the questionnaire. Please respond to each question as carefully as you can. Extra space is provided for you to write comments or elaborate on your answers, if you desire. All answers and comments will be confidential.

1. How many years have you been teaching in this school? ____
2. How many separate preparations do you have each day? ____
3. How many students do you teach over the whole day? ____
4. How many minutes of planning time are you allotted each day? ____

5. How often do teachers in your department:	Seldom or Never	Occasionally	Often	Very Often
a. Help each other develop tests?	___	___	___	___
b. Share course outlines with each other?	___	___	___	___
c. Observe each other's classes?	___	___	___	___
d. Discuss students' learning problems with each other?	___	___	___	___

6. Please indicate the extent of your agreement or disagreement with each of the following statements:	Strongly Agree	Agree	Disagree	Strongly Disagree
a. I make heavy academic demands on students.	___	___	___	___
b. School has a special responsibility to students who are failing.	___	___	___	___
c. There is too much pressure on teachers to assess and evaluate students' learning.	___	___	___	___
d. If I didn't have to grade students, I would rarely if ever give tests.	___	___	___	___
e. I strictly enforce work deadlines in my courses.	___	___	___	___

Comments ____

PLEASE ANSWER ALL REMAINING QUESTIONS IN TERMS OF THIS COURSE: ____________

7. How many years have you been teaching this course? ____
8. Is this course taught by other teachers at your school? Yes__ No__

IF 'YES,' ANSWER QUESTIONS 9-10; IF 'NO,' SKIP TO QUESTION 11.

9. Do you and other teachers of this course:
   a. Use the same text? Yes__ No__ Don't know __
   b. Use the same course outlines and schedules? Yes__ No__ Don't know __
   c. Use the same tests? Yes__ No__ Don't know __

10. How often do you talk to other teachers of this course about your:	Seldom or Never	Occasionally	Often	Very Often
a. Course goals?	___	___	___	___
b. Course tests?	___	___	___	___
c. Grading standards?	___	___	___	___

11. How important to you is it that all sections of this course meet the same learning objectives?

Not important__ Somewhat important__ Very important__

12. When you grade students in this course at the end of the term, how often do you:	Seldom or Never	Occasionally	Often	Very Often
a. Grade according to a curve?	___	___	___	___
b. Allow students to raise their grade by completing makeup assignments?	___	___	___	___
c. Allow students to raise their grade by doing optional extra-credit assignments?	___	___	___	___
d. Require students to bring their work up to a preset standard before you assign a passing grade?	___	___	___	___
e. Give a student who has achieved important learning goals, or who has shown a lot of progress, a higher grade than scores on tests or other work strictly would allow?	___	___	___	___

Comments ____

WHICH CLASS PERIOD WILL YOU HAVE COMPLETE THE STUDENT QUESTIONNAIRE? ________
PLEASE ANSWER ALL REMAINING QUESTIONS IN TERMS OF HOW YOU TEACH THIS CLASS.

1. What percentage of students in this class:

	5% or less	About 10-20%	About 25-35%	More than 35%
a. Is not ready for the level of work?	___	___	___	___
b. Is too advanced for the level of work?	___	___	___	___

14. Did you give a test at the beginning of the year to assess student readiness for this class? Yes___ No___

15. Will you give a final exam in this class? Yes___ No___

If "yes," what percentage of the student's grade will come from the final exam? 10% or less___ 15-20%___ 25-30%___ More than 30%___

16. In this class, about how often do you give:

	Once in 2 weeks or less often	Once a week	Several times a week	Every day
a. Short quizzes?	___	___	___	___
b. Homework?	___	___	___	___
c. Inclass worksheets?	___	___	___	___

17. In this class, how often do you give written assignments that require several days for the student to complete—for example, problem sets, laboratory reports, essays, stories, or book reports?

Less than once a month ___ About once every 3 weeks ___
About once a month ___ Once or more every 2 weeks ___

What kinds of written assignments do you give?

____

____

____

18. A teacher may decide to give separate subscores to different parts of such written assignments—for example, scores for different areas of learning such as knowledge and skills.

	Rarely if ever	Sometimes	Regularly
How often do you give separate scores to parts of such written assign...s in this class?	___	___	___

What aspects of learn... or work do you score separately?

____

____

____

WE NOW ASK YOU TO THINK ABOUT THE UNIT TESTS YOU GIVE IN THIS CLASS.

All remaining questions are about tests that deal with substantial sections of the course content. Such tests are often called "unit tests," and that term is used here, but we also mean tests on chapters and m... ...ading assignments. We don't mean quick quizzes, on one hand, or the final exa... ...the other. Please describe the unit tests you give in this class.

____

____

____

____

____

19. How often do you give unit tests in this class?

About once every week ___
About once every two weeks ___
About once every three weeks ___
About once every month ___
Less than once a month ___

20. What percentage of the student's grade for this class comes from unit tests?

Less than 20% ___
About 20-25% ___
About 30-35% ___
About 40-45% ___
50% or more ___

21. How often do you use the following methods to inform students in this class about what they will be expected to know on a unit test?

	Rarely if ever	Sometimes	Regularly
a. At the beginning of a unit or period of instruction, I inform students what they will be expected to know on the test for that unit or period.	___	___	___
b. During a unit or period of instruction, I give students samples of the kinds of questions to be included on the test for that unit or period.	___	___	___
c. A day or so before a test, I conduct a formal review of the knowledge and skills to be tested.	___	___	___

Comments ____

____

22. When you are constructing a unit test to be used in this class, how often do you use each of the following practices?

	Rarely if ever	Sometimes	Regularly
a. I seek advice from other teachers about the questions I've written.	___	___	___
b. I select or adapt test items that accompany textbooks or other curriculum materials.	___	___	___
c. I use a written list of student learning objectives as a guide for developing or selecting test items.	___	___	___
d. I include items that demand higher levels of thinking than classroom tasks.	___	___	___
e. I include items that require my judgment to score or evaluate, such as open-ended essay questions or problems that have more than one correct answer.	___	___	___
f. I establish standards for performance on a test that students must meet before they move on to new topics or units.	___	___	___
g. I develop alternate forms of tests in order to retest students who fail the first time around.	___	___	___

Comments ______________________________

____

23. When you are scoring unit tests in this class, how often do you use each of the following practices?

	Rarely if ever	Sometimes	Regularly
a. I give more weight or points to items that cover material stressed in class than to items that cover material I have not stressed.	___	___	___
b. I calculate separate scores for different sections of the test that deal with distinct learning goals—for example, knowledge of facts vs. application of skills.	___	___	___
c. I provide written comments on test papers to supplement or elaborate on test scores.	___	___	___
d. I discount particular items on a test that students show to be unfair or confusing.	___	___	___

Comments

	Rarely if ever	Sometimes	Regularly	N.A.
24. When you include items on unit tests in this class that require your judgment to score or evaluate, how often do you use explicit scoring guidelines, such as rating scales?	___	___	___	

What aspects of learning do you score separately?

____

____

25. When discussing results of unit tests with this class, how often do you:

	Rarely if ever	Sometimes	Regularly
a. Inform the class about particular items or sections of the test on which a large number of students did <u>poorly</u>.	___	___	___
b. Inform the class about particular items or sections of the test on which a large number of students did exceptionally <u>well</u>.	___	___	___

26. On how many unit tests in this class have:

	Few or none of the tests	About half the tests	Nearly every test
a. a large number of students done poorly?	___	___	___
b. a large number of students done exceptionally well?	___	___	___

27. When a <u>large</u> percentage of this class does <u>poorly</u> on a unit test, how often do you use the following practices:

	Rarely if ever	Sometimes	Regularly
a. After going over test results with the class, I spend one or more class periods reteaching the material.	___	___	___
b. A[illegible] going over test results with the [illegible]ss, I move on to the next topic or [illegible]it in order to keep to the course schedule.	___	___	___
c. I consult [illegible]her teachers about test content [illegible] student learning problems.	___	___	___

Comments ______________________________

____

BEST COPY AVAILABLE

28. When an individual student or small group of students in this class does poorly on a unit test, how often do you use the following practices?

	Rarely if ever	Sometimes	Regularly
a. I give special corrective or remedial instruction outside the regular class period.	___	___	___
b. I create special learning activities in class for such students in addition to regular class work.	___	___	___
c. I arrange for peer tutors.	___	___	___
d. I develop less difficult learning objectives for such students and adapt my teaching and my tests accordingly.	___	___	___
e. I talk to parents about students' learning problems.	___	___	___
f. I talk to other teachers about students' learning problems.	___	___	___
g. I refer students for special testing or help with learning problems.	___	___	___

Comments ______________________________

____

____

29. When an individual student or small group of students in this class does exceptionally well on a unit test, how often do you use the following practices?

	Rarely if ever	Sometimes	Regularly
a. I give special enriching or accelerating instruction outside the class period.	___	___	___
b. I ask such students to serve as peer tutors.	___	___	___
c. I create special learning activities in class for such students in addition to regular work.	___	___	___
d. I develop more advanced learning objectives for such students and adapt my teaching and my tests accordingly.	___	___	___

Comments ______________________________

____

____

30. Approximately what percentage of the points available on a unit test in this class comes from test items with a multiple-choice or short-answer format? ______

31. How many of the questions on your unit tests do you change from year to year?

Almost none____ Some____ Most____ Virtually all____

32. In order to understand your approach to testing and assessment of student progress in this class, what should we know that has not been covered by the items on this questionnaire? For example, has your department developed an explicit program of teaching and testing for this course, or have you adopted a particular strategy as a result of past training and experience? If you have time, we would appreciate your sharing your ideas with us on this page.

THANK YOU FOR COMPLETING THE QUESTIONNAIRE

APPENDIX B

Measures of Students' Academic Work and Participation in Class

The analysis of predictors of students' efforts to study is part of a larger research effort to build a model of predictors of students' academic effort and participation in schooling. Other analyses are underway using the present data base, and a concurrent study by Duckworth and deJung (1986) has identified influences on students' rates of skipping school and cutting class. In this appendix, we describe student response to seven measures of effort and participation that were shown in Table III-2 of the main text to be correlated with the Effort index (composed of questionnaire items on studying for tests and working hard in class). Table 1 of this appendix shows the distribution of student responses to those seven questions.

The first item (16) asked how many hours the student spent studying or doing homework for the class in which the questionnaire was administered. The data show that, except in geometry, a fourth or more of the students in each subject reported spending one hour at least a week. Never did a fourth of the students in a subject report studying more than three hours a week for the class. Biology students reported the least amount of time spent.

The second item (21) asked how often the student had homework done on time. Except in biology, half or more of the students in a subject responded "nearly always." Also except in biology, less than a fourth of the students in a subject reported only "occasionally" or "almost never." The first two items indicate that biology students in all reported expending less academic effort than students in the other three subjects. This was not inconsistent with the findings reported in Table II-1, which showed that biology and history students reported less often studying thoroughly before tests and that biology students reported less often working hard than English and history students, although not than geometry students.

The third item (24) asked about engagement in class rather than academic effort as such. In all, about a third of the students reported that their minds "nearly always" or "often" wandered during lessons in the class. Of all the students, biology students were most likely to respond "nearly always."

The fourth item (30) asked about participation in class pursuant to engagement. About a third of the students in all responded that they "nearly always" asked questions in class when they didn't understand. Biology students were least likely to respond "nearly always."

The fifth item (23) asked about a form of participation in class that required extra effort--seeking the teacher's help outside class time when a student didn't understand the material. Students were less likely to report taking this action "nearly always" than asking questions in class "nearly always," and more than a third of the students in each subject reported "almost never" seeking the teacher's help.

Finally, the sixth and seventh items (14 and 13) asked about avoidance of the class--temporary avoidance, in the form of coming to class late, and major avoidance, in the form of cutting class. As the table shows, about two thirds of the students in all reported hardly ever being late to class, and nearly 90 percent of the students in all reported hardly ever cutting class.

Reference: Duckworth, Kenneth, and deJung, John. "Variation in Student Skipping: A Study of Six Schools." Eugene, Oregon: Center for Educational Policy and Management, University of Oregon, 1986.

Appendix B

Table 1

Students' Academic Work and Participation

(Percentage of students, in each subject and in all, selecting questionnaire responses)

Questionnaire Item	Response	Biol	Geom	Engl	Hist	All
9. How many hours each week do you spend outside class studying or doing homework for this class *	> 5 hrs.	5	5	5	5	5
	4-5 hrs.	11	19	17	19	17
	2-3 hrs.	46	57	51	50	51
	1 hr. at most	38	20	27	26	27
21. I have my homework done when I come to class	Nearly always	43	54	52	50	50
	Often	31	28	28	30	29
	Occasionally	19	14	15	15	16
	Almost never	8	4	5	5	6
24. My mind wanders during lessons this class	Nearly always	18	12	13	15	14
	Often	15	20	17	20	18
	Occasionally	47	47	49	44	47
	Almost never	20	22	21	21	21
30. I ask questions in class when I don't understand	Nearly always	26	33	36	32	32
	Often	26	28	24	20	25
	Occasionally	32	24	26	29	28
	Almost never	15	16	14	19	16
23. [illegible] t under-sta[illegible]rial, I seek [illegible]acher's help ou[illegible]ide class time	Nearly always	13	15	19	10	14
	Often	17	16	16	14	16
	Occasionally	34	34	30	31	32
	Almost never	36	35	36	44	38
14. How often are you late to this class? *	More than once/wk	8	8	7	10	8
	Once a week	6	6	10	13	8
	Once every two weeks	14	14	15	18	15
	Hardly ever	71	71	68	60	68
13. How often would you say you have skipped or cut this class? *	Once a week or more	3	3	3	2	3
	Once every 2 weeks	2	4	4	6	4
	Once a month	4	6	5	6	5
	Hardly ever	92	87	89	85	88
Number of students		391	478	420	405	1694

* "Don't know" responses treated as missing responses.

** Actual number for each item depends on number of missing responses.