Pergamon

Int. J. Educ. Res. Vol. 23, No. 8, pp. 653–752, 1995
Copyright © 1996 Published by Elsevier Science Ltd
Printed in Great Britain. All rights reserved
0883–0355/96 $29.00

S0883–0355(96)00001–8

THE IMPACT OF CURRICULUM-BASED EXTERNAL EXAMINATIONS ON SCHOOL PRIORITIES AND STUDENT LEARNING

JOHN H. BISHOP

Center on the Educational Quality of the Workforce and Cornell Youth and Work Program and Center for Advanced Human Resource Studies, New York State School of Industrial and Labor Relations, Cornell University, Ithaca, NY 14853–0952, U.S.A.

CONTENTS

	PREFACE	655
	INTRODUCTION	657
CHAPTER 1	AN ECONOMIC THEORY OF THE EFFECTS OF PEER PRESSURE AND REWARDS FOR LEARNING ON SCHOOL PRIORITIES, STUDENT EFFORT AND ACHIEVEMENT	661
	Introduction	661
	The Achievement Function	661
	Choosing Learning Effort	663
	Determining Student Effort	665
	Determining Community Investment in School Quality	665
	The Effects of Pecuniary Rewards on Effort	666
CHAPTER 2	THE IMPACT OF THE PAYOFF TO COLLEGE AND THE SELECTIVITY OF COLLEGE ADMISSIONS ON EFFORT AND ACHIEVEMENT IN HIGH SCHOOL	669
	Introduction	669
	Evidence from Aggregate Time-Series Data	669
	Cross-Section Evidence	674
	Summary	676
CHAPTER 3	WHEN CURRICULUM-BASED EXAMINATIONS ARE ABSENT, WHAT HAPPENS TO ACADEMIC STANDARDS, INCENTIVES AND ACHIEVEMENT IN THE UNITED STATES?	679
	Systems of Curriculum-Based External Examinations — A Definition	679
	Contrasts with Other Types of Examinations	681
	The Effects of Curriculum-Based External Examinations on Student and Parent Behavior	683
	The Effects of Curriculum-Based External Examinations on Teacher Behavior	689
	The Effects of Curriculum-Based External Examinations on the Behavior of School Administrators and School Boards	691

CHAPTER 4 TESTING FOR BACKWASH EFFECTS OF EXAMINATION SYSTEMS: INTERNATIONAL COMPARISONS 695
Do Examination Systems Affect Teacher Salaries, Qualifications and Spending Per Pupil? 695
Impacts on Student Achievement in Math, Science and Geography: An Analysis of Data from the International Assessment of Educational Progress 697
Effects of Sweden's Elimination of Exit Examination 699

CHAPTER 5 NEW YORK STATE VS. THE REST OF THE UNITED STATES 705
Introduction 705
Does New York State Invest More in K-12 Education? 705
Do New York Schools Add More Value than the Schools of Other States? 707
The Impact of Curriculum-Based Exams on SAT Scores 707

CHAPTER 6 MATHEMATICS AND SCIENCE ACHIEVEMENT IN CANADA: COMPARING PROVINCES 711
Introduction 711
The IAEP Data 713
Impacts of Curriculum-Based Examinations on Math and Science Achievement 713
Other Determinants of Mathematics and Science Achievement 716
The Impact of Exam Systems on Parent, Student, Teacher and Administrator Behavior 720
How Does the U.S.A. Differ from Canada? 723

CHAPTER 7 CASE STUDY OF ENGLAND, SCOTLAND, FRANCE, THE NETHERLANDS AND THE UNITED STATES 725
Introduction 725
Differentials in Academic Achievement 725
The Proximate Causes of Achievement Differentials 727
Signaling as Ultimate Cause: External Examinations as Standard Setters 736

CHAPTER 8 SUMMARY AND LESSONS 745
Suggestions for Research 747
References 748

PREFACE

The preparation of this monograph was made possible by support from the Pew Charitable Trust, the German Marshall Fund, the Center for Advanced Human Resource Studies and the Center on the Educational Quality of the Workforce (agreement number R117Q00011–91, as administered by the Office of Educational Research and Improvement, U.S. Department of Education). The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

INTRODUCTION

Many countries have recently implemented changes in their curriculum-based examination systems. The English have merged the old Certificate of Secondary Education (CSE) and the O level exams into the General Certificate of Secondary Education (GCSE). France has broadened the list of Baccalaureate Examinations to include numerous vocational specialities and has set a goal of 80% of the age cohort participating by the year 2000. The Brevet exam at the end of lower secondary school, which had been abolished in 1977, was reintroduced in 1986. "The reasons were that the results had been declining in the experience of many people . . ." (Kreeft, 1990, p. 6). The Canadian province of Manitoba has recently reestablished curriculum-based examinations that had been discontinued in the early 1970s. Education Ministers in a number of the German *Länder* "are demanding that *Abitur* requirements should be defined centrally in all *Länder* and that standards should be harmonized to make the level of attainment compatible all over Germany" (WENR, 1993, p. 6).

Proposals to create or strengthen curriculum-based exams are central to the policy debate in the United States. The Education Subcouncil of the Competitiveness Policy Council (a nonpartisan group appointed by the President and congressional leaders) recently concluded that:

> Working hard and achieving in school must count for them [students], too. And presently, high school students who plan to go on to college do not need to work hard and get good grades in order to achieve their goal. Except for the tiny percentage of kids who want to go to selective colleges, students know that, no matter how poor their grades, they will be able to find a college that will accept them. If most colleges continue to admit students who have done little work in high school, there is no reason to expect any change in student behavior.
>
> The vast majority of employers give exactly the same message to students going directly from high school to work: What you did in high school does not count. . . . Hard working kids do not . . . have an edge since few employers ever inquire about what courses a young applicant took or ask to see a transcript (March 1993, p. 30).

Their recommended solution was that "*external assessments be given to individual students at the secondary level and that the results should be a major but not exclusive factor qualifying for college and better jobs at better wages*" (1993, p. 30).

The Commission on Workforce Quality and Labor Market Efficiency advocated similar reforms:

> The business community should . . . show through their hiring and promotion decisions that academic achievements will be rewarded (1989, p. 9).
>
> National educational and employer associations should work together to develop easily understood transcripts, based on voluntary achievement testing programs, that assess student proficiency in a wide variety of academic and vocational areas (1989, p. 12).

The American Federation of Teachers advocates developing a system in which:

> Students are periodically tested on whether they're reaching the standards, and if they are not, the system responds with appropriate assistance and intervention. Until they meet the standards, they won't be able to

> graduate from high school or enter college. . . . The curriculums that teachers use are based on the standards and so are the assessments . . . professional development programs are focused on preparing teachers to help students meet the standards . . . and textbooks and other instructional materials are tailored to the content of the curriculum frameworks (AFT, 1995, p. 1–2).

The theory that lies behind these policy recommendations can be stated very simply. Learning is not a passive act; it requires the time and active engagement of the learner. Students have many other uses for their time, so learning is costly for them. The intensity of their investment in learning depends on a comparison of benefits — intrinsic and extrinsic rewards for learning — to costs. When the benefits of learning increase, student effort increases and learning increases. Parents, school administrators and teachers are also influenced by comparisons of the benefits of learning to the costs of focusing school resources and policies on academic achievement. When rewards for academic achievement grow for students and parents, pressures are generated that induce teachers, administrators and school boards to refocus their energies on raising academic achievement.

Systems of regional or national curriculum-based examinations in secondary school that signal levels of achievement in specific subjects to colleges and employers transform the incentive environment of students, parents, teachers and school administrators and this stimulates changes in priorities and behavior that in turn result in higher achievement in examined subjects.

The policy debate is filled with claims about the effects of exam systems, but there is little solid empirical research on how they effect school policies, teaching and student learning. This issue begins the process of filling this gap in our knowledge. It proceeds as follows. Chapter 1 provides a formal mathematical presentation of the simple economic theory of learning and school priority setting summarized in the two previous paragraphs. Readers who are uncomfortable with mathematical models may want to skim or skip over the mathematics. Such readers are encouraged, however, to read as much of the text in the chapter as possible and to look ahead at Chapter 3 where the theory is presented in an informal manner with many examples and illustrations from American schools.

The theory assumes that student effort and school priority setting respond positively to increases in the economic payoff to learning and to obtaining a college degree. Chapter 2 presents a number of tests of this assumption of the theory.

According to our theory, curriculum-based external examinations increase the pecuniary and nonpecuniary rewards for study and this, in turn, induces students to study harder at school to refocus priorities on academic achievement. Chapter 3 offers a nonmathematical explanation of the theory together with examples of its operation taken from American experience. Sixteen hypotheses about the impacts of curriculum-based external exams are specified, many of which are tested later in the issue.

The effects of external examination systems must be assessed by comparing education systems, not by comparing individuals, classrooms or schools. This limits the number of independent observations and reduces the power of formal statistical tests. The low power of individual tests will be overcome by developing multiple tests of hypotheses. International comparative data on spending and learning outcomes is used in Chapter 4 to test for impacts of curriculum-based external exams. As predicted, countries with end of school examination systems invest more in their secondary schools and achievement in examination subjects tends to be higher.

Chapter 5 compares New York, the only state in the U.S. with curriculum-based external exams, to the rest of the nation and again finds support for the hypothesis. Chapter 6 compares Canadian provinces with curriculum-based external examinations to provinces lacking such systems. Here again the students in provinces with exams outperform the students who live in

provinces without them. Schools in provinces with exit exams hire better trained specialist teachers, have better science facilities and schedule more class time for mathematics and science.

Chapter 7 applies the theory to the task of explaining and interpreting differences in student achievement among France, England, Scotland, the Netherlands and the United States. The lag of England and Scotland behind France and the Netherlands cannot be explained by differences in examination systems, so other causes are explored.

The issue concludes with a summary and suggestions for future research.

CHAPTER 1

AN ECONOMIC THEORY OF THE EFFECTS OF PEER PRESSURE AND REWARDS FOR LEARNING ON SCHOOL PRIORITIES, STUDENT EFFORT AND ACHIEVEMENT

Introduction

Much of the economic research on elementary and secondary education has employed a production function paradigm. Conventionally, test scores measuring academic achievement are the outputs, teachers are the labor input and students are goods in process.

This chapter points in a different direction. Schools are viewed as worker-managed organizations producing multiple outputs. In the classroom/school team production unit, students are as much workers as the teachers. Students are also consumers who choose which goals (output) to focus on and how much effort to put into each goal. The behavior of each of the system's actors (teachers, administrators, school board, students and parents) depends on the incentives facing them. The incentives, in turn, depend upon the cost and reliability of the information (signals) that are generated about the various outputs of the system.

In order to focus attention on the essence of the problem a very simple theoretical model containing just six equations is developed below. The six equations are: an achievement function, a rewards for achievement function, equations for the benefits and costs of student effort and for the benefits and costs of school inputs focused on academic achievement.

The Achievement Function

Learning is a change that takes place in a person. It occurs when an individual who is *ready and able* to learn, is offered an *opportunity* to learn and makes the *effort* to learn. All three elements are essential.

Learning readiness and ability — indexed by A — depends on prior learning, intelligence and family background. A is exogenous (i.e., determined outside the model).

While, in principle, every literate individual with access to a library has the opportunity to learn, schools and teachers have, in practice, a great deal of influence on what youngsters learn and at what pace. Educators determine what courses are required, which electives are offered, the topics covered, teaching methods, homework and paper assignments and classroom expectations. In our model, opportunity to learn is operationalized as school quality or QX. School quality is the product of two variables, one exogenous, the other endogenous (i.e., determined within the model). X is a composite index of the teacher behaviors, school inputs and policies fostering academic achievement that are controlled by educators and their political supervisors.

X is the teacher and/or school administrator's choice variable. Q is an exogenous efficiency index measuring how effective X, the chosen school inputs and policies, are at improving student achievement. Q can also be viewed as an index of the teaching technology available to the educators.

While facilitating learning is the primary purpose of American schools, other goals compete for school resources and administrative attention. Among the other goals are child care, improved health and nutrition, discouraging drug use, student self-esteem, community service, opportunities for physical exercise, equalizing opportunities and/or outcomes, achieving racial balance in school populations and community entertainment (e.g., band and interscholastic sports). Consequently, X should not be confused with per pupil expenditure. The model is designed to explain the priority that schools give to academic learning as a goal, not school spending.

The part of the learning equation controlled by the student is effort (E). Students choose which courses to take and how much effort to devote to each course. Classroom observational studies have found that the time engaged in learning is considerably less than the time available for learning and that time on task varies a great deal across students and across classrooms (Frederick, 1977; Frederick, Walberg & Rasher, 1979). Time devoted to homework also varies a great deal. In 1990, 33% of 17 year olds reported doing two or more hours of homework each night, while another 15% reported not being assigned homework or not doing the homework assigned (NCES, 1993, p. 352).

Just as important as the time devoted to learning is the intensity of the student's engagement in the process. Sizer (1984) concluded, "No more important finding has emerged from the inquiries of our study than that the American high school student, *as student*, is all too often docile, compliant, and without initiative" (p. 54).

Student–teacher ratios of 17 to 1 imply that, in the aggregate, students spend nearly 17 times as many hours learning as teachers spend teaching. Student time and engagement is, therefore, probably the most important input in the educational process.

All of these factors are summarized in the simple equation:

$$\text{Human Capital at the end of secondary school} = L = AE^{\alpha}(QX)^{\beta},\ \alpha+\beta<1 \qquad (1)$$

where E = student effort — an index of the time and psychic energy that students devote to learning,
X = a composite index of inputs and policies that foster academic achievement,
Q = efficiency factor for X, the School Inputs Index,
α = elasticity of human capital (L) with respect to effort, and
β = elasticity of human capital (L) with respect to school academic quality (QX).

I have assumed the human capital production function is Cobb-Douglas with $\alpha + \beta < 1$. This gives the model a number of realistic features.

(a) School quality and student effort interact positively. An improvement in teacher quality enhances the effect of greater student effort and vice versa.
(b) A 20% increase in both effort (E) and school quality (QX) increases human capital (L) by less than 20% .

The Extrinsic Rewards for Achievement Function

Present Discounted Value of Pecuniary and other External Rewards for Achievement =

$$Y = wL + \theta\,(L - L_m) + \lambda A \tag{2}$$

where w = the impact of absolute levels of achievement (human capital) at the end of secondary school on the present discounted value of lifetime earnings and other extrinsic rewards for learning. It includes the effects of secondary school learning that operate through admission to preferred colleges, completing college, successful pursuit of lucrative but difficult majors such as engineering, entry into and completion of graduate programs. It also includes the benefits that parents derive from the economic success of their children and the honor and prestige given to those who are seen as high achievers. Curriculum based external examinations increase w, the payoff to absolute achievement, and tend to reduce θ, the payoff to one's relative position (rank) in the secondary school's graduating class and λ, the payoff to IQ and family background.

L–L_m = student achievement relative to the school mean (L_m). Rank in class and grades awarded on a curve are examples of signals of achievement that describe the student's achievement relative to others in the school.

θ = the impact of achievement relative to the school mean on the present discounted value of lifetime earnings and other extrinsic rewards for learning (includes impacts on admission to preferred colleges and the benefits that parents derive from the success of their children).

λ = the impact of A (i.e., early IQ, early achievement and family background) on the present discounted value of lifetime earnings. λ would be large and w and θ small if access to college depended solely on family background and IQ test scores obtained prior to entering secondary school. The SAT is not a pure IQ test, but relative to curriculum-based exams it is at the aptitude end of the spectrum. Consequently, substituting curriculum-based exams for the SAT in university admissions decisions would lower λ and raise w. Since A is exogenous, changes in λ do not effect student incentives to study or community incentives to invest in secondary schools.

A key feature of this model is that studying by one student imposes costs on other students. Greater achievement for person 'i' increases school mean achievement, L_m, and lowers everyone else's position relative to the mean (e.g., rank in class). In fact the loss that others experience when person 'i' tries harder is exactly equal to $\theta(L–L_m)$, the gain person 'i' experiences from raising her achievement relative to the school mean. If, for example, there are 200 students in the graduating class, a one unit increase in L_i for person 'I' raises L_m by 0.005. This lowers the benefits that each of the 199 other students get from the second term of Equation (2) by 0.005*θ.

Choosing Learning Effort

The student is not just another input into an educational process that is controlled and directed by others. Rather, students are more appropriately viewed as entrepreneurs trying to grow and

develop in an environment shaped by a great variety of forces: history, parents, teachers, peers, employers, government, society, and their own abilities. Students generally, and secondary school students especially, should be viewed as choosing what to study and how hard to work at learning it.

When making these choices the student compares expected benefits to expected costs. The benefits are both extrinsic (Y) and intrinsic (j) — the joy of learning for its own sake and the honor and respect that parents and teachers give for achievement.

Student and Parent Benefits of Effort Equation

$$\text{Student and Parent Benefits of Effort} = B = (w + \theta + j)\left[AE^{\alpha}(QX)^{\beta}\right] \tag{3}$$

where j = the present discounted value of the intrinsic nonpecuniary benefits, *joy*, of learning received by the student and her parents. Note that these benefits are assumed to occur regardless of whether the learning is signaled to others or honored publicly.

Student and Parent Costs of Effort Equation

The costs are time, psychic energy, money for tuition and books, loss of control over one's in-class time and, frequently, peer pressure against learning.

$$\text{Costs of Student Effort} = C = C_0E^{\eta} + c_1\theta L \tag{4}$$

where C_0E^{η} = the costs that arise from giving up other more pleasurable activities when time and energy are devoted to learning.

η = the elasticity of this cost with respect to effort. $\eta > 1$ because the marginal costs of effort rise as effort increases.

$c_1\theta L$ = the costs of effort that result from the fear that one's classmates will think you are a "*nerd . . . teacher's pet . . . or acting white.*" When $c_1 = 1$, the anti-*nerd* pressure against academic effort exactly offsets the losses that trying harder imposes on others $\theta(L–L_m)$ because greater achievement for person 'I' increases school mean achievement, L_m, and lowers everyone else's position relative to the mean (e.g., rank in class). If $c_1 > 1$, anti-nerd peer pressure imposes larger costs on the studious than they impose on their classmates.

Community Benefits of Giving Academics High Priority Equation

Communities similarly base their choices about spending and priorities on a comparison of benefits and costs. Community decisions about school spending and the priority attached to learning are motivated by benefits that students get from general improvements in their academic achievement, w and j.

$$\text{Community Benefits of Giving Academic Achievement Top Priority} = B_p = p\,(w_m + j_m)\left[A^{m}E_m^{\alpha}(QX)^{\beta}\right] \tag{5}$$

where p = a parameter characterizing the political power of parents in school board elections and the benefits that nonparents get from greater achievement.

w_m= the average impact of absolute levels of achievement (human capital) on the extrinsic benefits of learning (includes impacts on admission to preferred colleges and the benefits that parents derive from the economic success of their children).

j_m = the average present discounted value of the nonpecuniary benefits, *joy*, of learning received by students and parents.

Note how θ, the benefit of improving one's ranking relative to school mean achievement, does not appear in the equation describing community benefits of greater learning.

Cost of School Inputs Equation

Cost of Schools Inputs and Policies Devoted to Academic Achievement

$$= D = D_0 X^{\gamma} \tag{6}$$

where γ = the elasticity of the costs of school inputs with respect to increases in X. $\gamma > 1$

Determining Student Effort

To study the determinants of student effort, we define a net benefits of study effort equation, $B - C$, by subtracting Equation (4) from Equation (3), and then obtain the maximum of the function by differentiating it with respect to E, effort.

$$\max B - C = (w + \theta + j)\left[AE^{\alpha}(QX)^{\beta}\right] - \left[C_0 E^{\eta} + c_1 \theta L\right] \tag{7}$$

The derivative of Equation (7) with respect to E for the average student is:

$$\frac{d(B-C)}{dE} = \alpha\,(w + \theta + j)\left[AE^{\alpha-1}(QX)^{\beta}\right] - \eta C_0 E^{\eta-1} - \alpha c_1 \theta\left[AE^{\alpha-1}(QX)^{\beta}\right] = 0 \tag{8}$$

The third term is combined with the first.

$$= \alpha\left[w + j + (1 - c_1)\theta)\right] * \left[AE^{\alpha-1}(QX)^{\beta}\right] - \eta C_0 E^{\eta-1} = 0 \tag{9}$$

$$E^{\eta-1}/E^{\alpha-1} = E^{\eta-\alpha} = \frac{\alpha}{\eta C_0}\left[w + j + (1 - c_1)\theta)\right] * \left[A(QX)^{\beta}\right] \tag{10}$$

$$E = \left\{\left[a/\eta C_0\right] * \left[w + j + (1 - c_1)\theta\right] * \left[A(QX)^{\beta}\right]\right\}^{1/(\eta-\alpha)} \tag{11}$$

Determining Community Investment in School Quality

We determine the community's investment in school quality by subtracting the costs of that investment, (Equation (6)), from its benefits (Equation (5)), and then determining the maximum for the function by differentiating with respect to X.

$$B_p - D = p\,(w_m + j_m)\left[AE_m\alpha(QX)^\beta\right] - D_0X^\gamma \tag{12}$$

$$\frac{dB_p - D}{dX} = \beta p\,(w_m + j_m)\left[AE_m{}^\alpha Q^\beta X^{\beta-1}\right] - \gamma D_0 X^{\gamma-1} = 0 \tag{13}$$

$$X = \left\{\left[\beta / \gamma D_0\right](pw_m + pj_m)\left[AE_m{}^\alpha Q^\beta\right]\right\}^{1/(\gamma-\beta)} \tag{14}$$

The Effects of Pecuniary Rewards on Effort

What does this model tell us will happen to effort, the school input index and achievement when curriculum-based external exams are introduced? Curriculum-based exams cause an increase in w, the extrinsic reward for absolute achievement. The effect of an increase in w can be determined by differentiating Equation (11) and Equation (14) with respect to w.

$$\frac{dE}{dw} = \frac{1}{(\eta - \alpha)[w + j + (1 - c_1)\theta]}\,E > 0 \tag{15}$$

$$\frac{dX}{dw} = \frac{1}{(\gamma - \beta)(w_m + j_m)}\,X > 0 \tag{16}$$

Employing the chain rule, we differentiate L with respect to w.

$$\frac{dL}{dw} = \frac{dL}{dE}\frac{dE}{dw} + \frac{dL}{dX}\frac{dX}{dw} = \frac{\alpha L}{(\eta - \alpha)[w + j + (1 - c_1)\theta]} + \frac{\beta L}{(\gamma - \beta)(w_m + j_m)} > 0 \tag{17}$$

The model clearly predicts that a rise in the economic payoff to the absolute level of human capital (L) at the end of high school from whatever cause — technical change, shifts in college admissions policies or establishing a system of curriculum-based examinations — will have substantial positive effects on student effort, student learning and inputs devoted to raising academic achievement. The impact of economic rewards is largest when nonpecuniary motives are less important than pecuniary motives (i.e., j is considerably smaller than w), when anti-nerd peer pressure is strong (e.g., $c_1 > 1$), when the elasticities of learning with respect to effort (α) and school inputs (β) are substantial, and when the marginal cost curves for E and X are flat (e.g., η and γ are close to 1).

Curriculum-based external exams (CBEEs) increase w, the pecuniary rewards for absolute levels of academic achievement, in two ways. First, they improve the signals of high school achievement available to colleges and employers and thus cause these institutions to give greater weight to absolute levels of academic achievement when they make admissions and hiring decisions. In the process their decisions become less sensitive to other factors such as aptitude tests like the SAT, family connections, racial and religious stereotypes, recommendations of previous employers and the chemistry of 20 minute interviews. Total extrinsic rewards for achievement will doubtless rise, $d(w+\theta)/d(\text{CBEE}) < 0$.

Second, curriculum-based external examinations can be expected to shift attention and rewards from measures of relative achievement such as rank in class and teacher grades to measures of absolute achievement (e.g., grades on the external exam). In mathematical terms, $dw/d(\text{CBEE}) > 0$ probably implies a $d\theta/d(\text{CBEE}) < 0$. The decline in θ does not influence school priority decisions because θ does not appear in the equations describing the benefits and costs of X, school inputs. The likely decline in rewards for relative achievement has an ambiguous

effect on student effort. The equation for this impact is:

$$\frac{dE}{d\theta} = \frac{(1-c_1)}{(\eta-\alpha)[w+j+(1-c_1)\theta]} E \lesseqgtr 0 \tag{18}$$

Since c_1 is positive, Equation (18) is always less positive than Equation (15). Thus, in the unlikely event that a curriculum-based exam decreased θ by the same amount it increased w, student effort would increase because peer pressure against studying would decrease. The positive c_1 reflects the fact that anti-nerd peer pressure against studying is caused in part by institutions that reward relative rather than absolute achievement. When one student's success comes at the expense of other students, students as a group pressure their classmates to be uncompetitive. If peer pressure is so strong that it outweighs the pecuniary benefits that students expect from being ranked at the top of the class, $c_1 > 1$ and a reduction in these rewards (θ) will increase student effort.

During the last 15 years the payoff to college attendance has more than doubled. The wage reward for mathematics achievement also doubled for males and increased 25% for females between 1978 and 1986. These facts imply that w increased significantly during this period. The model predicts that this should have caused teachers and administrators to place higher priority on fostering academic achievement and induced students to study harder and learn more. These hypotheses are tested in Chapter 2.

Chapter 3 provides a thorough discussion of why curriculum-based external examinations increase extrinsic payoffs for academic achievement and thus induce changes in the behavior of students, parents and educators. These hypotheses are then tested in Chapters 4 through 6.

CHAPTER 2

THE IMPACT OF THE PAYOFF TO COLLEGE AND THE SELECTIVITY OF COLLEGE ADMISSIONS ON EFFORT AND ACHIEVEMENT IN HIGH SCHOOL

Introduction

It is well established that the pecuniary payoff to college has substantial effects on the proportion of high school graduates who enter and complete college (Freeman, 1975, 1976b; Bishop, 1977, 1990). Studies have also found that the choice of undergraduate and graduate fields of study responds rapidly to changes in the labor market payoff to different fields of study (Freeman, 1971, 1976a). What impacts do the payoff to college have on study effort and achievement in high school?

Academic achievement in high school has important effects on the probability of attending college, the quality of college attended and the probability of completing college. Consequently, the selectivity of college admissions decisions and the payoff to the college degree are primary determinants of w, the extrinsic rewards to studying in high school. For those not planning to go to college, better jobs are the relevant pecuniary inducement for study in high school. The theory presented in Chapter 1, therefore, predicts that the dramatic rise during the 1980s in the payoff to college and to mathematics achievement of high school graduates who did not go to college should have stimulated schools to set higher standards and induced students to study harder. This in turn should have increased achievement. These hypotheses are tested in time series and cross-section data in this chapter.

Evidence from Aggregate Time-Series Data

In this section we will review how the rewards for learning in high school have evolved during the last 60 years. Three indicators of these rewards are examined:

- The wage payoff to academic achievement for those who do not go to college.
- The impact of academic achievement on the probability of college attendance.
- The payoff to completing a bachelor's degree.

Trends in the Payoff to Academic Achievement Holding Years of Schooling Constant

In the United States, reading and mathematics skills have historically not had large effects on the wage rates of young workers who have not gone to college (Bishop, 1985). Even so, the

payoffs to these skills appeared to decline during late 1960s and early 1970s. The threat of litigation brought under the 1971 Griggs interpretation of Title 7 of the Civil Rights Act of 1965 induced many employers to stop using tests assessing reading and mathematics to help in the selection of new employees (Friedman & Williams, 1982). As a result, during the 1970s young high school graduates who had learned English, science and mathematics thoroughly, typically did not earn appreciably more than the high school graduates who had done poorly in these subjects (Bishop, 1985, 1989, 1993). Since individuals who have strong mathematics skills and good grades in high school are better employees (Hunter, 1983), their greater productivity is eventually recognized and rewarded by employers. It often takes more than a decade, however (Grubb, 1994; Bishop, 1989). Thus noncollege bound students who work hard in high school must wait many years to reap any rewards. When they do appear, the rewards — a 2–3% increase in earnings per grade level equivalent on achievement tests — are not a very strong incentive for hard study. Even for adults the wage rate rewards of learning are smaller than the actual gain in productivity that results (Bishop, 1987, 1993).

The environment changed somewhat during the 1980s and 1990s. The economy was being transformed by the spread of the personal computer and the introduction of high-performance work systems empowering front-line workers to take over functions previously performed by engineers and managers. Court actions such as the Wards Cove decision made it easier to defend using reading and mathematics tests as part of a selection process. A number of employers reintroduced basic skills tests into their selection procedures for clerical and factory jobs. A 1985 American Society for Personnel Administration survey (BNA, 1986) found that 24% of the firms responding had increased testing in the past year and another 44% were considering an increase in the amount of testing they do. Employers now appear to be paying more attention to the academic qualifications and, as a result, the labor market rewards for mathematical ability of young workers have risen. Murnane, Willett and Levy (1994) found that the effect of a one standard deviation increase in mathematics skills on the wage rates of 24 year old men rose from $0.46 per hour in 1978 to $1.15 per hour in 1986. The wage payoff for young women rose from $1.15 per hour to $1.42 per hour in 1986.

Trends in College Admissions Selectivity

During the 1920s need-based scholarships were rare and social class was the primary determinant of who went to college. Achievement in high school had little effect on who went to college. Those who entered college in 1925 were on average at the 53rd percentile of ability among high school graduates. Those who did not enter college were at the 47th percentile on average; a gap of only six points (Taubman & Wales, 1972).

In the decades that followed academic achievement in high school became a substantially more important determinant of college entry. The percentile gap between college entrants and high school graduates not going to college rose to 11 points in 1929, 15 points in 1934, 20 points in 1946, 19 points in 1950, 22 points in 1957, 27 points in 1960–61 (Taubman & Wales, 1972). By 1960 achievement had larger effects on college entry probabilities than social class (Bishop, 1977).

In order to examine what has happened since 1960, the calculations made by Taubman and Wales were replicated in two more recent nationally representative longitudinal studies of high school seniors — National Longitudinal Survey: Class of 72, and High School and Beyond (1980 graduates). Rates of college entrance for the year following graduation were calculated for each quartile of the ability distribution. The relationship between college entrance rates and a student's ability ranking was approximated by a series of linear segments and mean ability

rankings were calculated for college entrants and for noncollege going high school graduates.

The trend toward growing academic selectivity of college enrollment reversed during the 1960s and early 1970s. A large number of 2 year colleges with open admissions policies were founded during this period and males flooded into colleges to postpone being drafted into the military. The class rank gap between those who do and do not attend college fell from 27 percentile points in 1960–61 to 22 points in 1972 (calculated from data in Gardner, 1987; Bishop, 1991). The correlation between test scores and college attendance 18 months after graduating from high school fell from 0.458 for 1961 graduates to 0.399 for 1972 graduates (Bishop, 1991).

After 1972 the trend reversed once again. The class rank gap rose to 25 points by 1980 and the correlation between test scores and college attendance rose to 0.442 (Jackson, 1988). A 1985 survey of 2203 college admissions directors allows us to track admissions selectivity between 1980 and 1985. Only 2% of the institutions reported they had become less selective during that period. The proportion reporting they had become more selective, on the other hand, was 42% at four-year private colleges, 49% at four-year public colleges and 8% at two-year public colleges (Breland, Wilder & Robertson, 1986). Later surveys of college admissions officers indicate that the rigor of the courses taken in high school has become an increasingly important admissions criterion and reliance on the SAT has decreased.

In summary, the dependence of college entry on achievement in high school appears to have grown over the course of the last 70 years. It rose dramatically from 1925 to 1960, fell back for about 15 years and then resumed its rise in the late 1970s.

Trends in the Payoff to College

The wage premium received by those with a college education has varied a great deal over the course of the last 65 years. The ratio of the weekly earnings of five high-level professional occupations to the weekly earnings of manufacturing production workers fell substantially during the depression and World War II from 3.3 in 1929, to 2.88 in 1939 and 2.46 in 1950 (Bureau of the Census, 1975, D804, D913–D916, D920; Bishop, 1991). Professional wages then rebounded and the ratio rose to 2.9 in 1964.

Decennial census data on the payoff to college exhibit a similar pattern. The ratio of the incomes of 25–35 year old white males with four or more years of college to the incomes of 20–29 year old high school graduates was 1.99 in 1939, 1.45 in 1949, 1.68 in 1959 and 1.76 in 1969 (Marenbach, 1973, pp. 89–90, see Figure 2.1).

Figure 2.1 also presents Katz and Murphy's estimates of college/high school weekly earnings ratio for workers with less than 5 years of post-school work experience. Their carefully constructed series indicates that the wage premium for college degree holders with under 5 years of work experience fell from 51% in 1963–64 to 44% in 1979–80 and then rose to 84% in 1988. The payoff to obtaining an associates degree has also increased substantially during the 1980s and 1990s.

Test Score Trends — 1918–1980

Despite significant increases in the share of the age cohort in school, test scores of school children rose during the period between 1920 and 1940. A study of the school children of eastern Tennessee found that 7th and 8th graders improved by two-thirds of a standard deviation on paper and pencil IQ tests between 1930 and 1940 (Wheeler, 1942). A study of two high

schools in the midwest found no change in the mean IQ test score of the students at a small rural high school and a five-point increase between 1923 and 1942 at a large high school serving a small city and the surrounding county (Finch, 1946). Johnson (1935) found a three-point gain in IQ test scores between 1925 and 1935 at Grover Cleveland High School in St. Louis. Roessel's (1937) comparison of the students in three Minnesota high schools in 1920 and 1934 and Rundquist's (1936) comparison of Minneapolis high school students in 1929 and 1934 both found increases in test scores. World War II draftees had IQ test scores more than two-thirds of a standard deviation higher than World War I draftees (Tuddenham, 1948; Flynn, 1984).

For the post-WWII era, the best data on trends in the academic achievement of students nearing completion of secondary schooling comes from the Iowa Test of Educational Development (ITED). This data set is particularly valuable because it provides equated data extending back to 1942 and annual data from 1960 to the present. Because about 95% of the public and private schools in the state of Iowa regularly participated in the testing program, the analysis of trends in ITED data for Iowa is not plagued by changing selectivity of the population taking the test. This feature of the data makes ITED trends for Iowa a better representation of national trends prior to 1970 than the ACT, the SAT, and the American Council on Education Psychological Exam. These other tests were taken at first by a highly selected group of students and only more recently by more representative samples of college bound students, so trends in scores on these tests are biased by the decreasing selectivity of those who took the test.

Figure 2.1 plots the trends of ITED composite scores for Iowa 11th and 12th graders. Through 1965 the trend was up: at first moderately so, and then dramatically after Sputnik (Forsyth, 1987; Hieronymus, Lindquist & Hoover, 1979). The gains for 12th graders between 1942 and 1966 are all the more remarkable for they coincide with an increase in the high school graduation rate in Iowa from 65% in 1941 to 88% in 1968. Other tests that have been administered for long spans of time to stable test-taking populations also exhibit a positive trend during this period (Farr & Fey, 1982). Between 1958 and 1966 Minnesota high school juniors gained 0.39

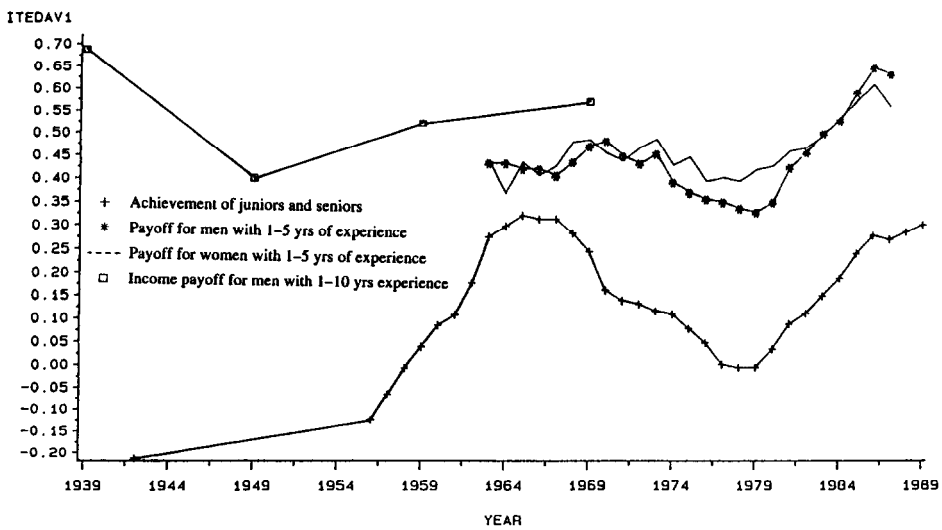

Figure 2.1. The payoff to college and academic achievement of high school students.

SDs on the Minnesota Scholastic Aptitude Test (Swanson, 1973). The periodic national standardizations of the ITED also exhibited an increase during the 1960s (Koretz, 1986).

In 1966 the educational achievement of high school students stopped rising and began a decline that lasted about 13 years. On the ITED the composite scores of Iowa 9th graders dropped 0.283 SDs and the scores of seniors dropped 0.35 SDs or about 1.25 grade level equivalents. Comparable declines occurred throughout the country and for upper elementary and junior high school students as well. The bottom was reached around 1980.

The Response of Effort to the Rising Payoff to Learning Since 1980

The increase in rewards for achievement during the last 15 years appears to have helped cause a rise in standards and in student study effort. Many states established minimum competency tests for graduation and increased the number of mathematics and science courses required for graduation. Homework assigned and completed has increased. The percentage of 13 year olds reporting they either had no homework or did not do it fell from 33% in 1982 to 9% in 1990. The percentage of 17 year olds reporting they did at least one hour of homework each day rose from 32.5% in 1978 to 66% in 1990 (NCES, 1993, pp. 122, 351). High school graduation rates are creeping up. Students are taking more rigorous courses. Between 1982 and 1990 enrolment shares rose 20 percentage points for Geometry, 14 points for Algebra II, 19 points for Chemistry and 8 points for Physics (NCES, 1993, p. 68). The number of students taking AP exams has tripled.

Test Score Trends Since 1980

Around 1980 the test scores of Iowa high school students began to rise again. By 1988 Iowa 12th graders had recouped about three-quarters of their previous decline and 9th graders had surpassed their 1965 record by almost two-fifths of a grade level equivalent. SAT and ACT scores have risen as well although at a slower rate because of increases during the 1980s in the proportion of high school graduates taking these tests.

Between 1982 and 1992 NAEP mathematics and science scores at age 17 rose more than a grade level equivalent. Grade level equivalents on the NAEP IRT scale scores can be calculated by dividing the difference between the scores of 17 year old students and 13 year old students by four. Using this simple approach we can see that a GLE is 8.5 points on the NAEP mathematics scale, 9 points on the NAEP science scale and 7.5 points on the NAEP reading scale (Mullis *et al.*, 1994). Mathematics scores of 17 year olds rose 1.06 GLEs (9 points) and science scores rose 1.22 GLEs (11 points). Reading scores have risen 0.53 GLEs since 1980. These gains have been accomplished despite significant growth in the share of students who are from minority and disadvantaged backgrounds. Between 1977 and 1990, the black and Hispanic share of NAEP test takers rose from 16 to 23%. The achievement of minority students improved by roughly 2 grade level equivalents between 1980–82 and 1992.

Comparing Trends

The time-series data for the postwar period appear to suggest that changes in the economic payoff to college may influence the academic achievement of students in high school. Both the

college payoff and test scores rose during the 1950s and peaked during the 1960s. The two series then decline together and starting in 1979–80 they start up together. When the average 11th and 12th grade ITED test score is regressed on the logarithm of the current college graduate/ high school graduate weekly earnings ratio and a trend, one obtains the following results:

$$\text{ITEDAV} = \underset{(5.56)}{22.07} + \underset{(7.92)}{1.41}*\log(\text{college Wage/NSWage} - \underset{(3.70)}{0.0114}*\text{Year}) \qquad R^2 = 0.78 \quad 1966\text{–}1987$$

The results suggest that the decline in the payoff to college between 1969 and 1979 lowered test scores by 15% of a standard deviation (about half the total decline in academic achievement at the end of high school) and the rise between 1979 and 1987 raised test scores by 29% of a standard deviation. Further support for the hypothesis comes from the fact that categories of students who have higher than average probabilities of attending college — white students, suburban students and college going students — experienced larger than average test score declines between 1966 and 1979 (Koretz, 1986).

If the model is expanded to incorporate changes in college selectivity, it can also explain rising test scores during the 1920s and 1930s. On the surface, the declining payoff to college during that period appears to be inconsistent with our theory. However, the growing impact of high school achievement on college attendance means that the expected payoff to such achievement (payoff times the impact of a one standard deviation test score differential on college entrance) rose substantially. Using the wage premium of professionals over manufacturing workers as our indicator of the payoff to a college degree, the expected payoff to a one SD differential in achievement was 12.4% in 1924, 22.7% in 1929, 29.1% in 1939 and 24.9% in 1949. Only the 1940s creates problems for the theory. During the 1940s test scores rose despite declines in the payoff to college and expected payoff to learning in high school.

There may, however, be other explanations for the post-1966 test score decline and rebound during the 1980s. Graduation requirements and teacher expectations appear to have followed the same cycle. While the theory explains these changes in expectations and school policies as responses to shifts in economic payoffs, others might argue that the causes lie elsewhere. As a result, the examination of aggregate time series data for just one country can never be completely decisive evidence for our hypothesis or indeed any hypothesis. The data provide just too few degrees of freedom. Other types of data need to be examined. We turn now to tests of the economic payoff hypothesis in cross-section data.

Cross-Section Evidence

Cross-section data provide a second opportunity to examine how the college payoff influences the behavior of high school students. There is spatial variation in the future payoffs to college education so, if most young people intend to remain in the local labor market after school, one would expect geographic differentials in the payoff to affect (a) the number of college prep courses taken in high school, (b) the time spent studying and (c) the probability of attending college. In previous work I have found that a rather crude measure of the college payoff — the average differential between an accountant's, teacher's and engineer's wages and operative wages — had significant positive effects on the college attendance rates of most students in the top 75% of the ability distribution (Bishop, 1977).

Table 2.1 presents linear regression estimates of the effect of college payoff, academic orienta-

Table 2.1
Determinants of College Entrance

Independent Variable	Poverty	Lower middle income	Middle income	Upper middle income	High income
College payoff (SD = $570)	0.080*** (2.86)	0.054*** (3.44)	0.036*** (2.89)	−0.033* (1.83)	−0.018 (1.14)
Academic orientation	0.163*** (7.58)	0.201*** (16.16)	0.256*** (24.33)	0.227*** (14.54)	0.227*** (15.87)
Hours of Study (SD = 5.5 hrs)	0.092*** (4.59)	0.043*** (3.96)	0.116*** (12.64)	0.057*** (4.16)	−0.022 (.12)
R^2	0.224	0.273	0.307	0.288	0.272
Number of observations	2320	6538	8766	4309	5113

Standardized regression coefficients with t statistics in parentheses under the coefficient. The coefficient on the payoff variable represents the effect of a $570 increase in the difference between professionals and operatives at a time (1959) when mean earnings of male high school graduates 25–64 years old was $6132. The coefficient on the study hours per week variable represents the effect of a 5.5 hour increase in reported study time. Mean hours of reported study time including study halls was about 9 hours.
***$P < 0.01$ on a two-tail test.
**$P < 0.05$ on a two-tail test.
*$P < 0.10$ on a two-tail test.
Source: Weighted least squares prediction of college attendance in fall 1961 using longitudinal data on 27,046 male high school juniors in the Project Talent database. Students were categorized by family income and separate models were estimated for each group. An extensive set of controls was included in the models: socioeconomic status, number of siblings, the number of changes of school, academic aptitude, the tuition at public universities and colleges in the state, the cost (including travel costs) of attending the lowest cost 2 year and 4 year colleges, distance to the lowest cost college, the selectivity of local colleges, the opportunity cost of the student's time (the operative wage rate), and a dummy for being from an intact family and the cheapest local postsecondary institution is a 2 year vocational college. Data were collected by phone from a 5% sample of the nonrespondents to Project Talent's mail questionnaires. Because nonrespondents to the mail questionnaire were systematically different from those who responded, the people who were part of the nonrespondent sample were assigned weights of 20 in the weighted regression.

tion of courses and study time on the subsequent college attendance of 27,046 high school juniors at Project Talent high schools in 1960. Separate models were estimated for students categorized by family income. The control variables included in the regression are listed at the bottom of the table. Academic orientation of courses had substantial positive effects on college attendance and hours of study had modest positive effects.

The payoff variable for this analysis was the earnings differential between professional workers and operatives measured in 1959 dollars deflated for the local cost of living. The local labor market was either the SMSA of residence or the non-SMSA portion of the state. In 1959 male high school graduates 25–64 years old earned an average of $6132 and the payoff variable had a mean of $2957 and a standard deviation of $570. The college payoff variable had significant direct effects on college attendance rates of students from low and moderate income families even when hours spent studying, the academic orientation of courses and aptitude were controlled. Since students probably base judgements about the reward to college on both local and national data, these results are probably a lower bound estimate of the aggregate effect of a nationwide change in the payoff to college on rates of college attendance.

Do, however, prospective payoffs to college influence behavior of students while they are in

high school? To explore this issue, the measure of the college payoff employed in the college attendance regression was included in models predicting the number of college prep courses taken through the junior year of high school and the weekly number of hours spent studying including in-school study periods. The standardized regression coefficients (representing the effect of a 20% change in the payoff to college) are presented in Table 2.2. The findings are that students living in labor markets with a large college payoff took additional academic courses but they did not spend more time studying. The absence of the expected positive effect of payoff on hours studying may be due to the inclusion in the study time variable of in school study periods, for college bound students typically take heavier course loads and consequently schedule fewer study periods.

Summary

The overall results are consistent with the hypothesis that economic returns to learning influence course taking and achievement in secondary school. The decline in the payoff to college

Table 2.2
The Effect of College Payoff on The Academic Orientation of High School Courses and Hours of Study Time

	Beta coefficient on payoff variable by income group				
	Poverty	Lower middle income	middle income	Upper middle income	High income
Regression predicting academic orientation	0.103*** (4.01)	0.025* (1.81)	0.011 (0.98)	0.065*** (4.26)	0.072*** (5.19)
Regression predicting study time	–0.060** (2.14)	–0.002 (0.10)	–0.057*** (4.33)	–0.081*** (4.54)	0.004 (0.25)
Number of observations	2320	6538	8766	4309	5113

Standardized regression coefficients representing the effect of a $570 (in 1959 dollars) increase in the earnings differential between professionals and operatives. The payoff variable had a mean of $2957 and a standard deviation of $570. Male high school graduates 25–64 years old earned an average of $6132 in 1959 ($t$ statistics are in parentheses under the coefficient).
***$P < 0.01$ on a two tail test.
**$P < 0.05$ on a two tail test.
*$P < 0.10$ on a two tail test.
Source: Weighted least squares models predicting the academic orientation of course taken and time spent in study halls and studying at home using data on 27,046 male high school juniors in the Project Talent database. Students were categorized by family income and separate models were estimated for each group. An extensive set of controls was included in the models: socioeconomic status, parents' education, academic aptitude, religious activity, the tuition at public universities and colleges in the state, the cost (including travel costs) of attending the lowest cost 2 year and 4 year colleges, distance to the lowest cost college, the selectivity of local colleges, the opportunity cost of the student's time (the operative wage rate and the SMSA unemployment rate), and characteristics of the local high school — size, teacher salary, teacher experience, homogeneous grouping, hours of homework assigned and dummies for race, being the eldest child, being from an intact family and sports ability. Data were collected by phone from a 5% sample of the nonrespondents to Project Talent's mail questionnaires. Because nonrespondents to the mail questionnaire were systematically different from those who responded, the people who were part of the nonrespondent sample were assigned weights of 20 in the weighted regression.

and to academic achievement apparently contributed to the decline in test scores during the late 1960s and 1970s. The subsequent increase in college selectivity, the payoff to college and to math achievement for high school graduates probably contributed to the rise in test scores during the 1980s.

CHAPTER 3

WHEN CURRICULUM-BASED EXAMINATIONS ARE ABSENT, WHAT HAPPENS TO ACADEMIC STANDARDS, INCENTIVES AND ACHIEVEMENT IN THE UNITED STATES?

Systems of Curriculum-Based External Examinations — A Definition

Let us begin by defining what is meant by a system of curriculum-based external examinations (CBEE). To have the effects that the Competitiveness Policy Council and the American Federation of Teachers have envisioned, a national or state system of examinations must have all nine of the following characteristics. It must:

(1) *Produce signals of student accomplishment that have real consequences for the student.* The nature of the rewards for achievement on these exams varies from country to country. In many systems exam results are averaged with teacher assessments to generate final grades for a specific course. In some cases passing the exam is necessary for graduation from lower or upper secondary school and/or confers the right to enroll in university. Exam results also often influence which university or academic program secondary school graduates can be admitted to, and hiring decisions of employers.

(2) *Define achievement relative to an external standard, not relative to other students in the classroom or the school.* Fair comparisons of achievement across schools and across students at different schools are now possible. This feature makes exam grades useful for selection decisions made by universities and employers. It also makes exam grades useful indicators of average achievement levels at a school. Parents making choices about where to live and where to send their child to school will now have something other than the socio-economic status of the neighborhood upon which to base that decision. Costrell's (1994a, b) formal analysis of the optimal setting of educational standards concluded that more centralized standard setting (state or national achievement exams) results in higher standards, higher achievement and higher social welfare than decentralized standard setting (i.e., teacher grading or schools graduation requirements).

(3) *Organized around specific disciplines.* This focuses responsibility for preparing the student for particular exams on one (or a small group of) teacher/s. When exams are tied to particular courses with specific content, the choice of courses effectively determines which examinations the student takes. Where students in different tracks or specialities study a subject for varying amounts of time, separate examinations are often given to the students in the different tracks. Indeed, requiring everyone to take the same exam appears to force a lowering of the passing standard.

(4) *Signal multiple levels of achievement in the subject.* If only a pass–fail signal is gener-

ated by an exam, the standard will have to be set low enough to allow almost everyone to pass and this will not stimulate the great bulk of students to greater effort. If the single standard were significantly higher, it would be too high for many to reach and large numbers of students would become completely discouraged.

(5) *Assess a major portion of what students studying the subject are expected to know and be able to do.* The content domain from which the questions and assigned tasks are sampled must represent important material that we want students to learn. It is not, however, essential that the exam assess every dimension of achievement we care about. Teachers can be given responsibility for evaluating the dimensions of performance that cannot reliably be assessed by external means.

(6) *Be perceived as fair.* All students need to do the exam under the same conditions, be asked the same questions (or offered the same selection of questions) and have the same amount of time to complete the exam. The requirement of test security means that different questions must be devised each time the exam is administered. When good rubrics for grading the exam have been developed in advance, the student's own teacher can be one of the graders of the paper. In some countries, papers are graded centrally by committees of teachers who do not know the candidates. A perception of fairness is essential because examination results are going to have important consequences.

(7) *Reliably measure achievement.* If an exam is an unreliable measure of achievement in the subject, the grades received by students will have a large random component and study effort will have limited impacts on examination grades. Students will eventually discover this and the positive incentive effects of improved signals of student achievement will be lost. If exam grades are unreliable, the fairness of allocating prized places in universities on the basis of exam grades will be called into question and political support for the high-stakes exam will be undermined.

(8) *Evaluate mastery of a particular curriculum that is taught in schools.* The examination must be aligned with the curriculum taught. This maximizes the perception that effort will be rewarded.

(9) *Cover a majority of secondary school students.* Exams for only a small elite segment of an educational system influence standards in that segment but will probably have limited effects on the bulk of students. Since all countries allow secondary school students to specialize, this will generally require a diverse set of exams, not a single exam taken by all. Individual examinations may target special groups of students (e.g., those studying Russian or auto mechanics), but the system must cover the majority of students.

Curriculum-based examinations are generally tied to taking a particular sequence of courses and are therefore effectively voluntary. They cover a broad range of subjects and the questions used are typically published after the test has been given.

The Advanced Placement (AP) examinations are the one exception to the generalization that the U.S. lacks national curriculum-based external examinations. Students who take these courses and pass the examinations may receive college credit for high school work. While it is growing rapidly, AP is still a very small program. In 1991 only 9,786 of the 22,902 U.S. high schools offered any AP courses. Only 37 AP exams were taken on average in each participating high school. Of the graduating class in 1993, only 3.2% had taken an AP English exam, 4.1% an AP history exam, 3.8% an AP calculus exam, and 3.7% an AP science exam (NCES, 1993; National Education Goals Panel, 1993). Low participation prevents the AP exams from being a CBEE system all by itself. It can, however, serve as a component of a larger system.

Of the 50 states, only New York State has a curriculum-based examination system covering

just about everyone. New York's minimum competency tests, the RCTs, set a minimum standard that everyone must pass. AP exams now effectively set standards for the top 5 or 10% of students — those seeking admission to highly selective colleges. Regents examinations for more than 25 different high school courses set the standard for the rest of the college going students.

Speaking of the system as it existed in 1966, Assistant Commissioner for Examinations and Scholarships, Sherman Tinkelman, said:

> The Regents examinations are closely related to the curriculum in New York State. They are, as you can see, inseparably intertwined. One supports and reinforces the other. . . . These instruments presuppose and define standards. . . . They are a strong supervisory and instructional tool — and deliberately so. They are effective in stimulating good teaching and good learning practices (Tinkelman, 1966, p. 12).

Sponsorship by the state Board of Regents is crucial to the role these examinations have played in setting and maintaining high standards and promoting reform. On occasions, examinations have been deliberately revised to induce changes in curriculum and teaching.

> For years our foreign language specialists went up and down the State beating the drums for curriculum reform in modern language teaching, for change in emphasis from formal grammar to conversation skills and reading skills. There was not very great impact until we introduced, after notice and with numerous sample exercises, oral comprehension and reading comprehension into our Regents examinations. Promptly thereafter, most schools adopted the new curricular objectives (Tinkelman, 1966, p. 12).

Contrasts with Other Types of Examinations

Having described what a CBEE is, let us now review what it is not. CBEEs may include authentic tasks and problems, but this is not essential. Authenticity is a desirable trait of a CBEE, but it is not a defining trait.

CBEEs often use extended answer-essay formats. Students are often asked to write essays, to show work while solving multistep mathematics problems, conduct scientific experiments or make something. But this too is not a defining trait. Certain skills and types of achievement are more reliably and cost-effectively measured using a multiple choice format. Consequently, most modern CBEEs use a mix of formats.

CBEEs generally provide criterion referenced interpretations of examination results. This is a desirable feature, but not a defining one. Norm referenced tests may be able to serve a CBEE function if the norm is national or regional and stable over time.

Commercially prepared achievement tests such as the CAT, CTBS, ITBS, ITED that are often mandated by state departments of education are not CBEEs because they fail test 1. Students have no stake in doing well in the exam. Where there are stakes attached to student performance, it is teachers and school administrators who experience the consequences, not individual students. Another problem with these tests is the lack of security. There are only a few forms of each test and revisions are infrequent. Different districts administer them at different times of the year. If real stakes became attached to individual performance on such tests, bootleg copies would become available and some students would study by memorizing the answers for the three different forms of the test. Consequently, these standardized achievement tests cannot be considered CBEEs because they fail requirements 6 and 1.

The SAT-I reasoning tests are not curriculum-based external exams because they do not meet tests 3, 5 and 8. They are not organized around school subjects and fail to assess most of the material — history, science, economics, civics, literature, foreign languages and the ability

to write an essay — that high school students are expected to learn. The SAT was designed from the beginning to minimize backwash effects on teaching and student study habits. Indeed, Richard Gummere, Harvard College's admissions director at the time the machine scored multiple-choice Scholastic Aptitude Test (SAT) replaced the curriculum-based essay style College Board Examinations, was very candid about why the SAT had been adopted:

> Learning in itself has ceased to be the main factor [in college admissions] The aptitude of the pupil is now the leading consideration (Gummere, 1943, p. 5).

The subject-specific SAT-II achievement tests sponsored by the Educational Testing Service meet some of the requirements of a CBEE. They are external, secure, fair, reliable, organized by subject and signal multiple levels of achievement. The stakes attached to test results are not very high, however. Colleges admit on the basis of the SAT-I, not the SAT-II achievement tests. The tests are limited to a multiple choice format and this calls into question whether they meet requirement 5 and 8 that a major portion of what students are supposed to learn in a subject is assessed. The final problem with the SAT-IIs is the relatively small number of students taking them, requirement 9.

From its inception, the ACT test was developed as a broad-spectrum achievement test, not an aptitude test. It fulfills CBEE requirements 1, 2, 4, 6, 7 and 9. It has subtests covering science and history so it is much broader in its curricular coverage than the SAT-I. It is not, however, assessing the mastery of a specific math, history or science curriculum as the Regents exams do, so it fails requirement 8. Within the time limitation of three hours it does a good job, but it lacks an assessment of writing, English literature and foreign languages. Since time is so limited, the reliability and validity of the history and science subtests are probably not high. Most educators probably feel that the ACT fails requirement 5. While it fails to meet all nine requirements, it comes close. A plausible argument could be made that the ACT test should have some of the same incentive effects that have been hypothesized for CBEEs, although in a weaker form.

The minimum competency exams that many American states require students to pass in order to graduate from secondary school are not CBEEs because they fail requirements 3 and 4. These minimum competency tests focus on basic skills that are taught in primary school and lower secondary school. While minimum competency exams have apparently reduced the numbers of students with very low basic skills levels (Lerner, 1991), the level of competence required to pass most of these exams is not very high. The tests are typically first taken in 9th or 10th grade and most students pass on the first sitting. High school transcripts indicate only whether the student eventually passes the test, not achievement levels above the minimum. Thus, for the great majority of students who pass them on the first try, the tests no longer stimulate study. Incentive effects are focused on the small minority who fail on the first try and must repeat the test. Minimum competency exams can be a useful part of a CBEE system, but other much more demanding curriculum-based exams signaling much higher levels of performance are even more critical.

The requirement (4) that a CBEE signals different levels of achievement (not just whether the student has achieved a minimum) is extremely important for it has major effects on the incentive effects of the exam. By age 13 students differ dramatically in their levels of achievement. On the National Assessment of Educational Progress, 7–9% of 13 year olds are four or more grade level equivalents behind their age mates and 15–17% are four or more grade level equivalents ahead of their age mates. When achievement differentials among students are as large as this, incentives for effort are stronger for most students if the full range of achievement is signaled rather than just whether the individual has passed some absolute standard. When

only a pass–fail signal is generated by a test, many students pass the standard without exertion and are, therefore, not stimulated to greater effort by the reward given for passing. Some of the least well-prepared students will judge the effort required to achieve the standard to be too great and the benefits too small to warrant the effort. They give up on the idea of meeting the standard. Only a few students will find the reward for exceeding the single absolute cutoff an incentive for greater effort (Kang, 1985). Costrell agrees: "The case for perfect information [making scores on external examinations available rather than just whether the individual passed or failed] would appear to be strong, if not airtight: for most plausible degrees of heterogeneity, egalitarianism, and pooling under decentralization, perfect information not only raises GDP, but also social welfare" (1994a, p. 970).

The Effects of Curriculum-Based External Examinations on Student and Parent Behavior

Now that we have defined what a curriculum-based external examination is and is not, we can proceed to describe its likely impact on students, parents, teachers, and school administrators. We begin with impacts on student and parents.

External assessments of achievement in specific secondary school subjects that have real consequences for the student increase rewards for learning and this should induce the student to choose more demanding courses and work harder in them.

Impacts on the Labor Market Rewards for Learning

Many American students see little connection between how much mathematics and science they learn and their future success in the labor market. When asked whether they agree with "It is important to know mathematics to get a good job," 32% of American 10th graders either disagreed or said they were "not sure." When the same question was asked about science, 21.5% disagreed and 35.4% were not sure (LSAY, 1992, Q# GB32K & GB33K).

The skepticism about the career utility of mathematics and science is, to some degree, justified. While additional schooling had large effects on wages during the early 1980s, when years of schooling is held constant, greater competence in science, reading and mathematical reasoning was associated with *lower* (not higher) wages and higher unemployment for young male high school graduates. Only technical competence and speed in doing simple arithmetic had significant positive effects on wages and earnings of men who had recently graduated from secondary school (Bishop, 1989b, 1993). For young women, science competence had no effect on earnings but verbal ability, mathematical reasoning and computational speed had modest effects on wages and earnings. A one grade level increase in all three of these abilities increased wages of young women by 1.5% and earnings by 3.1%.

Does this imply that academic achievement in secondary schools does not make young males more productive workers? No it does not. The grades awarded by teachers in German secondary schools are good predictors of scores three years in the future on apprenticeship exams with a heavy practical component. Industrial psychologists have conducted hundreds of studies, involving hundreds of thousands of workers, on the relationship between productivity in particular jobs and various predictors of that productivity. They have found that competence in reading, mathematics and problem solving are strongly related to productivity in almost all of the civilian and military jobs studied (Ghiselli, 1973; Hunter, Crossen & Friedman, 1985).

Additionally, American employers apparently believe that performance in school is correlated with performance at work. When asked to rate job applicant résumés containing information on grades in high school, they gave significantly better ratings to applicants with high grade point averages (Hollenbeck & Smith, 1984). Fifteen percent of small- and medium-sized employers ask job applicants to report high school grades on the job application (National Federation of Independent Business, 1987).

Why then the discrepancy between the productivity effects of academic achievement and its wage rewards? The first problem is that getting reliable information on academic achievement is generally difficult and costly. Self-reported grades may not be a reliable indicator of actual grades. American secondary schools are not very good at responding to employer requests for student transcripts, even when job applicants have given written permission. In 1987, high school transcripts had been obtained by 14.2% small- and medium-sized firms hiring high school graduates (NFIB, 1987). Grading standards vary a great deal across schools, however, so grades are more a signal of student achievement relative to others at the school than a signal of absolute achievement levels.

A system of state or national curriculum-based exams would remedy the problem of non-comparable standards and grading relative to others in the school. It would generate a set of reliable signals of absolute levels of achievement in high school that could be cheaply obtained by colleges and employers. If the legal environment was permissive, it would be reasonable to expect employers to start using exam results in their hiring decisions and that would then generate stronger and more immediate pecuniary rewards for learning in high school.

The absence of curriculum-based exams is probably not, however, the most important reason why rewards for academic achievement in secondary school are so weak in the United States. Employer administered employment tests are also seldom used. Such tests had been given to only 2.9% of the high school graduates hired in 1987 by small- and medium-sized firms (NFIB, 1987). The primary barrier to giving absolute levels of academic achievement greater weight in employer hiring decisions is the fear of legal challenges to such practices. After the 1971 Griggs decision, almost all firms were forced to stop employment testing by Equal Employment Opportunity Commission guidelines that made it prohibitively costly to demonstrate test validity (Friedman & Williams, 1982).

The situation is very different in Europe and Japan. In many European countries, job applications ask students to report their final grades in examination subjects (Raffe, 1984). Exhibit 1 is an example of one such application. Note how the applicant includes information on O and A level grades even though he has graduated from a polytechnic college. Résumés also conventionally include this information. Rosenbaum's (1990) study of the transition from high school to work in Japan found that good grades, no discipline problems, and participation in extracurricular activities all had significant positive effects on obtaining jobs at large firms and entering a white collar (rather than a blue collar) occupation. In the U.S., by contrast, the job outcomes of males were not improved by good grades, participation in extracurricular activities, fewer absences from school and a lack of discipline problems. For female high school graduates, obtaining a white collar job was associated with high grades, but it was also positively associated with being a discipline problem in school.

Two hypotheses are suggested by this discussion:

H.1 — By improving the reliability and validity of signals of achievement in high school and making them available to employers at no cost, an examination system will encourage employers to pay more attention to academic achievements when they hire recent graduates of secondary school. The impact of examination systems on rewards for learn-

APPLICATION FOR AN APPOINTMENT HANDLED BY MVP

16, Highfield Road, Edgbaston, Birmingham, B15 3DU Tel: 021 455 9765/0559

3318

Appointment applied for DISTRIBUTION PROJECTS MANAGER / RECRUITMENT CONSULTANTS Ref.No. 'FOCUS APRIL 89'

PERSONAL DETAILS: (block capitals)

Surname Title MR Forenames

Address TOP FLAT,

Postal Code SL5 0DA Tel.No.Home Work N/A

Marital Status SINGLE Children/Dependants (with ages) none

Age 28 Date of Birth 21/01/61 Nationality BRITISH Place of Birth BROMLEY, KENT

State of health EXCELLENT Height 5'6" Weight 12 STONE

Any disabilities/recurrent medical problems? NONE Regd.disabled N/A

Driving Licences FULL Car Owner NO Company Car YES

Endorsements, convictions, accidents, etc. 3 PTS – SPEEDING

Leisure activities and offices held in clubs and societies Skiing, Rugby (WINDSOR RUGBY CLUB), FOOTBALL, TENNIS, HILL WALKING, READING

EDUCATION:
Secondary Education

From	To	School	Exams Taken (inc. grades)	Other achievements
1974	1980	WINDSOR GRAMMAR	A levels – GEOG(E), HIST(D) Economics(O) O levels – Maths(B), English(B), ART(A) Geog(B), History(A), Biology(C)	1ST XV RUGBY 1ST VII RUGBY Lightning Chess Champ. Prefect, Art finalist Tennis finalists (Doubles). ADVANCED Skiing.

Further Education

From	To	College/University	Course & results (inc.class/grades)	Other achievements
1980	83	Plymouth Poly.	B.Sc Geography (HONS) – 2.2	1ST XV Rugby 1ST V Basketball. Course FOOTBALL, RUGBY.

Other training and qualifications (inc. in-company and external courses, etc.)

From	To	Establishment	Training/Qualifications
		Divons Stores Group	In-house – Time management. – Training skill – General management techniques.

Membership of professional bodies

Date	Association/Institute	Grade of membership	Offices held

Exhibit 1. Student job application.

ing will depend on whether employers are permitted to use examination grades when they screen job applicants.

H.2 — If the existence of an examination system increases the proportion of employers that use grades in examinations subjects to screen job applicants, students should be more likely to report that an examination subject is important for getting a good job.

Impacts on Enrollment in Demanding Courses

When curriculum-based exams are absent, schools are induced to offer and many students are induced to sign up for courses that have the reputation of being fun and not requiring much work to get a good grade. As one student who had avoided the harder courses even though she was sure she could do the work explained her decision: "*Why should I do it*, [the extra work], *if I don't have to?*" (Ward, 1994). As Theodore Sizer has observed, "A lot of the honors students aren't questers. They dodge the hard problems, the hard courses, to keep their averages up" (1985, p. 53).

Most parents are uninformed about course options and their consequences and often fail to influence the choices made. In the university town of Ithaca, New York, for example, less than one-fifth of parents attend the meeting in 8th grade at which the student and guidance counselor plan the student's 9th through 12th grade course sequence.

Guidance counselors are not encouraged to push extra students into the more demanding classes. Parents who want their child in the more demanding course are accommodated, but, behind their back, they are referred to as "pushy" or "elitist." For the great majority of students with nonpushy parents, most counselors see their role as insuring that the pupil does not get in over his head. No one should risk failure.

In many cases the parents of children assigned to less demanding courses are not informed that their child might have taken a more demanding course. In Ithaca, for example, 6th graders learn which of four types of mathematics classes they have been assigned to on the first day of middle school. If parents were to look at the schedule their child brings home they might see the word "enriched" and conclude that their child is taking the most demanding course. In fact, their child is in a class that moves at a slower pace than the "accelerated" class. If calculus is to be taken in senior year, a switch into the accelerated stream is essential, but it becomes increasingly difficult as time passes. Only a few parents are aware of the long-term consequences of the 6th grade math class assignment and the middle school makes no effort to inform parents of these consequences early enough to arrange a transfer into "accelerated" before the accelerated students have covered a good deal of ground not covered in the "enriched" class.

Many students manipulate the information their parents receive, so as to minimize the academic demands that are placed on them. This is illustrated by the following story:

> *In my case I stalled many attempts of my parents to get involved in my education. I wanted to be left alone in high school, and the high school was more than willing to accommodate my wish. Any parental influence could have damaged the treaty, and, therefore, I was cognizant that I needed to keep my parents out of my high school relationship. My parents' lack of information about the high school was my biggest ally in accomplishing my objective. They knew that I took math, science, and English, but in a modern high school those titles are meaningless as you could sleep through most classes and still receive above average grades. The other factor in my advantage was that my parents did not know the level of my scholastic ability. It was difficult, if not impossible, for them to know if my average to above average grades were good or bad since grades are a very relative measure of performance. For some students an average grade is great, while for others it is poor.*
>
> *My high school offered a variety of classes to deal with the diversity of its population. If asked, its administrators and faculty would claim that it has devised the perfect curriculum to meet the vast demands of the com-*

munity. If that is true, how do you explain my situation. At fourteen years of age my high school gave me the freedom to choose my destiny. At that time I chose to mentally opt out of school. My decision not only went unchallenged, it was furthered with the help of the system and teachers that allowed free choice at the expense of my education. The system created social norms that encouraged my apathy and under-achievement. I was so entrenched in the culture of the unspecial that I made conscious efforts to keep my parents from interfering with my decisions. It was not hard to do as the system was set up to give the illusion that everyone was receiving an equal and credible education (Jeff, 1991, 7–8).

Subjects are taught at vastly different levels, but the rigor of these courses is not well signaled to parents, relatives, neighbors, employers and colleges. Employers do not pay attention to the rigor of the courses taken in high school. Admissions staff at selective colleges learn how to read the transcripts of high schools they recruit heavily from and they evaluate grades in that light. However, at the beginning of the 1980s, most colleges did not factor the rigor of high school courses into their admissions decisions. Students who did not aspire to attend a selective college, consequently, quite rationally avoided rigorous courses and demanding teachers. Most students chose courses that had reputations of being fun and not requiring much work to get a good grade. Teachers knew this and adjusted their style of teaching and their homework assignments with an eye to maintaining enrollment levels:

An angry math teacher [who remembering] the elimination of a carefully planned program in technical mathematics for vocational students simply because not enough signed up for it, . . . [said]"Its easy to see who really makes decisions about what schools teach: the kids do (Powell, Farrar & Cohen, 1985, p. 9)."

The behavior of college admissions officers has started to change. Surveys by the National Association of College Admission Counselors (1993) indicate that they are increasing the weight they attach to taking rigorous courses in high school and doing well in these courses. Grades in high school have always been the number one consideration. Standardized test scores have now become the second most important consideration displacing class rank. Class rank is becoming less important because an increasing number of high schools are refusing to calculate class rank.

Our next hypothesis, therefore, is:

H.3 — Education systems with external examination systems will have more students pursuing rigorous courses of study.

Impacts on Peer Pressure

External assessments also have pervasive effects on the *structure* of student rewards. In their absence, signals of achievement tend to assess performance relative to other students in a class (e.g., grades and class rank). This gives students a personal interest in persuading each other not to study. The studious are called nerds, in part, because they are making it more difficult for others to get good grades or to be ranked near the top of the class. Since devoting time to studying for an exam is costly, the welfare of the entire class is maximized if no one studies for exams which are graded on a curve. The cooperative solution is "no one studies more than the minimum." Participants are generally able to tell who has broken the "minimize studying" code and reward those who conform and punish those who do not. Side payments and punishments are made in a currency of friendship, respect and ridicule that is not limited in supply. For most students the benefits that might result from studying for the exam are less important than the very certain costs of being considered a "brain geek," "grade grubber" or "acting White," so most students abide by the "minimize studying" "don't raise your hand too much" norm. The norm that seems to prevail is: *It is OK to be smart. You cannot help that. But, it is*

definitely not OK to spend a lot of time studying. Instead, use your free time to socialize, do sports or earn money. This is illustrated by the following story:

> *Erroneously I was lumped into the brains genus by others at [high] school just because of the classes I was in. This really irked me; not only was I not an athelete but I was also thought of as one of those "brain geeks." Being a brain really did have a stigma attached to it. Sometimes during a free period I would sit and listen to all the brains talk about how much they hated school work and how they never studied and I had to bite my lip to keep from laughing out aloud. I knew they were lying, and they knew they were lying too. I think that a lot of brains hung around together only because their fear of social isolation was greater than their petty rivalries. I think that my two friends who were brains liked me because I was almost on their level but I was not competitive* (Tim, 1986).

Note how those who broke the "minimize studying" norm tried to hide the fact from classmates. They did not espouse an alternative "learning is fun and important" norm.

The costs and benefits of studying vary across students because interest in the subject varies, ability varies and parental pressure and rewards vary. This heterogeneity means that some students break the "minimize studying" norm. When they are a small minority, they cannot avoid feeling denigrated by classmates. In the top track and at schools where many students aspire to attend competitive colleges, they are numerous enough to create a subculture of their own with norms denigrating those who do poorly on tests or who disrupt classroom activities. This is the structural basis of the "brains" and "preppie" cliques found in most American high schools. Most high school students, however, are in cliques that denigrate studying. At some school awards ceremonies, "some in the crowd jeer 'Nerd!'" as students are called to come up to receive an award (Suskind, 5/26/94, p. 1).

Peer pressure not to study does not derive from laziness. In jobs after school and at football practice, American adolescents work very hard. In these environments they are part of a team where individual efforts are visible and appreciated by teammates. Competition and rivalry are not absent, but they are offset by shared goals, shared successes and *external measures of achievement* (i.e., satisfied customers or winning the game). On the sports field, there is no greater sin than giving up, even when the score is hopelessly one sided. On the job, tasks not done by one worker will generally have to be completed by another. For too many students in too many American high schools, when it comes to academics, there is no greater sin than trying hard.

When learning is assessed relative to an outside standard, students no longer have a personal interest in getting the teacher off track or persuading each other to refrain from studying. Consequently, our theory predicts that:

H.4 — Educational systems employing external assessment will have less peer pressure against studying and fewer class disruptions intended to get the class off track. Impacts should be greatest in the years preceding the exam.

H.5 — External examinations will result in higher achievement, even when student characteristics, school resources, curriculum, teacher qualifications and teaching techniques are held constant. The effects should be strongest in 11th and 12th grade, but they should reach down to lower grades as well.

But the effects should not stop there. External exams also transform the incentives faced by the adults in the system. I would expect parents to set stricter limits on the amount of TV their children can watch and to be more likely to talk to their child about what they are learning in examination subjects. Our next set of hypotheses, therefore, are that:

H.6 — External exams will result in students spending less time watching TV.

H.7 — External exams will induce parents to spend more time talking with their children

about school and result in students perceiving their parents to be more interested in their doing well in examination subjects.

When external examinations are absent, parents tend to pressure teachers for less homework and better grades. In one school:

> Students were given class time to read *The Scarlet Letter, The Red Badge of Courage, Huckleberry Finn*, and *The Great Gatsby* because many would not read the books if they were assigned as homework. Parents had complained that such homework was excessive. Pressure from them might even bring the teaching of the books to a halt. . . . [As one teacher put it] "If you can't get them to read at home, you do the next best thing. It has to be done. . . . I'm trying to be optimistic and say we're building up their expectations in school (Powell, Farrar & Cohen, 1985, p. 81)."

Therefore:

H.8 — External exams will result in "parents no longer complain[ing] about too much homework or teachers who are too strict." They will "support teachers' efforts to elicit hard work from their children (AFT, 1995, p. 1)."

Opponents of external exams argue that focusing student attention on extrinsic rewards for learning will weaken students' intrinsic motivation to learn. Students will avoid opportunities to learn material that is not likely to be on the exam. George Madeus's list of negative effects includes "test scores come to be regarded by parents and students as the main, if not the sole, objective of education" and the result is "undue attention to material that is covered in the examinations, thereby excluding from teaching and learning many worthwhile educational objectives and experiences (1991b, p. 7)."

If they are right, students in systems with external exams should be less likely to read for pleasure or watch science programs like NOVA and Nature. Therefore, hypothesis 9 is that:

H.9 — Students will spend

- less time watching science documentaries on TV and
- less time reading for fun.

The Effects of Curriculum-Based External Examinations on Teacher Behavior

Impacts on Student–Teacher–Parent Team Building

External assessment of accomplishment puts students, teacher and parents on the same team. They assist the development of mentoring relationships between teachers and students. In the absence of external assessment, the effort to become friends with one's students and their parents tends to deteriorate into extravagant praise for mediocre accomplishment. In courts of law, judges must disqualify themselves when a friend comes before the bar. Yet, American teachers are placed in this double bind every day. Often the role conflict is resolved by lowering expectations or hiding failure with charitable phrases such as "does good work when he chooses to participate." Other times the choice of high standards means that close supportive relationships are sacrificed.

This is one of the considerations that has led most European secondary school teachers to support externally graded exams and external reviews of a student's completed projects. When changes in this system were proposed in Ireland, the Association of Secondary Teachers of Ireland wrote:

> *Major strengths of the Irish educational system have been:*
> *(i) The pastoral contribution of teachers in relation to their pupils.*

(ii) the perception of the teacher by the pupil as an advocate in terms of nationally certified examinations rather than as a judge.
The introduction of school-based assessment by the pupil's own teacher for certification purposes would undermine those two roles, to the detriment of all concerned. . . .
The role of the teacher as judge rather than advocate may lead to legal accountability in terms of marks awarded for certification purposes. This would automatically result in a distancing between the teacher, the pupil and the parent. It also opens the door to possible distortion of the results in response to either parental pressure or to pressure emanating from competition among local schools for pupils (ASTI, 1990, p. 1).

Therefore:

H.10 — External exams improve relationships between teachers and students. Teachers will shed the role of judge and take on the role of coach. Students will now see their teachers as allies in a common endeavor.

Impacts on Expectations and Grading Standards

Teachers are inevitably held somewhat accountable for how many of their students pass their courses. In systems without external exams, a teacher can lower class failure rates by lowering the passing standard. Thirty percent of American teachers say they "feel pressure to give higher grades than students' work deserves" and "feel pressure to reduce the difficulty and amount of work you assign" (Peter D. Hart Research Associates, 1994). Most commonly the pressure to lower standards comes from school administrators.

Students also pressure teachers to set low standards. Theodore Sizer's description of Ms. Shiffe's biology class illustrates the difficulties that some teachers get into:

She wanted the students to know these names. They did not want to know them and were not going to learn them. Apparently no outside threat — flunking, for example — affected the students. Shiffe did her thing, the students chattered on, even in the presence of a visitor. . . . Their common front of uninterest probably made examinations moot. Shiffe could not flunk them all, and, if their performance was uniformly shoddy, she would have to pass them all. Her desperation was as obvious as the students' cruelty toward her (1984, p. 157–158).

Theodore Sizer's (1984) description of Mr. Brody's class provides an example of how teachers benefit from setting modest goals.

He signaled to the students what the minima, the few questions for a test, were; all tenth and eleventh-graders could master these with absurdly little difficulty. The youngsters picked up the signal and kept their part of the bargain by being friendly and orderly. They did not push Brody, and he did not push them. . . . Brody's room was quiet, and his students liked him. No wonder he had the esteem of the principal who valued orderliness and good rapport between students and staff. Brody and his class had agreement, all right, agreement that reduced the efforts of both students and teacher to an irreducible and pathetic minimum (p. 156).

Some exceptional teachers are able to induce students outside the honors track to undertake tough learning tasks. But too often academic demands are compromised because the bulk of the class sees no need to accept them as reasonable and legitimate.

Under a system of external exams, teachers and local school administrators lose this option. Their response will be to strive to prepare their students for the external exam. And because of the exam, they will find they have more cooperative students. Therefore, hypothesis 11 is that:

H.11 — External exams will induce teachers

(a) to set higher standards,
(b) to assign more homework,
(c) to increase the number of experiments that students do in science class,
(d) to have students solve mathematics problems alone rather than in groups,
(e) to give more quizzes and tests,

(f) to increase their use of other teaching strategies which they believe improve exam performance.

H.12 — Impacts on how teachers teach will grow as the year of the exam approaches.

Some educators argue that external exams can have negative effects on teaching. It is argued, for example, that "preparation for high stakes tests often emphasizes rote memorization and cramming of students and drill and practice teaching methods" and that "some kinds of teaching to the test permit students to do well in examinations without recourse to higher levels of cognitive activity (Madeus, 1991, p. 7–8)."

The assumption of opponents appears to be that the tests developed by individual teachers for use in their class are better than examinations developed by the committees of teachers that would have responsibility for developing state or national examinations. To the contrary, the tests that teachers presently develop for themselves are generally of very low quality. Fleming and Chambers' (1983) study of tests developed by high school teachers using Bloom's taxonomy of instructional objectives found that "over all grades, 80% of the items on teachers' tests were constructed to tap the lowest of the taxonomic categories, knowledge (of terms, facts or principles)" (Thomas, 1991, p. 14). Rowher and Thomas (1987) found that in colleges fully 99% of items on instructor-developed tests in American history required the integration of ideas, while only 18% of junior high school and 14% of senior high school test items required such integration. Teachers test low-level competencies because that is what they teach. Students do not take state mandated tests in history, so poor history teaching cannot be blamed on standardized tests. More evidence is needed on this issue, so tests will be conducted of the following hypotheses:

H.13 — External exams provide fewer opportunities for exhibiting higher order thinking skills than conventional teacher-made tests.

H.14 — External exams will cause teachers to focus on teaching facts and definitions, not the scientific process. Students will conduct fewer experiments in science class and computation will be stressed in mathematics.

The Effects of Curriculum-Based External Examinations on the Behavior of School Administrators and School Boards

The behavior of school administrators and local school boards is also influenced by how student achievement is signaled to others.

Impacts on a School's Academic Standards

In the absence of valid measures of their students' achievement relative to an external standard, some American school administrators focus on lowering the failure rate rather than raising achievement. Teachers who set expectations too high get in trouble. For example, Adele Jones was fired because too many students (42% one year and 27% the next) failed her Algebra II course. The principal of the school justified his decision with the following:

> "*I have made it very clear that one of my goals is to decrease the failure rate, to make sure the kids feel good about learning, stay in class, stay in school and do well. . . . Math is just a big body of knowledge; what is Algebra II across the nation anyway?*" he asks. When he taught band, he adds, he certainly didn't expect kids to finish the year as musicians – but he did want them to know more about music than they did before. . . . All the talk about preparing students for college struck him as "*ludicrous*." Instead the goal should be to keep students studying math (Bradley, Sept 19, 1993, pp. 19, 20).

Senior Norman Kennedy said, however, the students who flunked Ms Jones class, "*were sleeping. They don't want to learn. They goof off, and they talk.*" At the hearing Walter Hall Jr., a student who had flunked the course, testified:

> "*I guess some of it could be attributed to a lack of study, because I wasn't really like into the books hour after hour, but in the rest of my classes, I was doing fairly well, and it was only testing that gave me a problem.*" He added that his parents had wondered how he could be getting such good grades in most classes without studying (p. 20).

Impacts on Hiring Standards and Resources Devoted to Academics

In the U.S. locally-elected school boards and the administrators they hire make the thousands of decisions that determine academic expectations and program quality (e.g., homework guidelines, whether to retain a popular but not very effective teacher, whether to raise wages to attract better teachers, etc.). Academic achievement is not the only goal of American schools. They are also expected to foster self-esteem, to provide counseling, supervised extracurricular activities, musical training, health services, community entertainment (e.g., interscholastic sports), drivers education and to do all this in a racially integrated setting. These other goals require additional staff and different kinds of staff. They may not be served by hiring teachers with a strong background in calculus or chemistry.

When there is no external assessment of academic achievement, students and their parents benefit little from administrative decisions that opt for higher standards, more qualified teachers or a heavier student work load. The immediate consequences of such decisions — higher taxes, more homework, having to repeat courses, lower GPAs, complaining parents, a greater risk of being denied a diploma — are all negative.

When student learning is not assessed externally, the positive effects of choosing academic rigor are negligible and postponed. If college admission decisions are based on rank in class, GPA and aptitude tests, not externally-assessed achievement in secondary school courses, upgraded standards will not improve the college admission prospects of next year's graduates. Graduates will probably do better in difficult college courses and will be more likely to get a degree, but that benefit is uncertain and far in the future. Maybe over time the school's reputation and, with it, the admission prospects of graduates will improve because the current graduates are more successful in local colleges. That, however, is even more uncertain and postponed. Publishing data on proportions of students meeting state targets on standardized tests probably speeds the process by which real improvements in a school's performance influence its local reputation. However, other indicators such as SAT test scores, proportions going to various types of colleges and the socioeconomic background of the students tend to be more prominent. As a result, school reputations are determined largely by things that teachers and administrators have little control over: the socioeconomic status of the student body and the proportion of graduates going to college.

Few American employers pay attention to a student's achievement in high school or the school's reputation when they make hiring selections (Bishop, 1989, 1993; Hollenbeck & Smith, 1984). Those that do pay attention to achievement use indicators of relative performance such as GPA and rank in class rather than results on an external exam as a hiring criterion. Consequently, higher standards do not benefit students as a group, so parents as a group have little incentive to lobby strongly for higher teacher salaries, higher standards and higher school taxes.

The Scholastic Aptitude Test is no substitute for curriculum based exams because it does not

assess knowledge and understanding of science, history, social science, statistics and calculus or the ability to write (Jencks & Crouse, 1982). Consequently, parents can see that improving the teaching of these subjects will have only minor effects on how their children do on the SAT, so why worry about standards?

External exams in secondary school subjects can be expected to transform the signaling environment. There is now a visible payoff to hiring better teachers and improving the school's science laboratories. Larger numbers of students pass the external exams and this in turn influences college admissions decisions. School reputations will now tend to reflect student academic performance rather than the family background of the community or the success of football and basketball teams. Hypothesis 15, therefore, is that:

H.15 — External exams will cause priorities to shift in favor of achievement in examination subjects. Administrators and school boards will be induced:

(a) to improve the school's science laboratories (if science is an examination subject),
(b) to offer additional courses in examination subjects and scale back offerings outside the core academic program,
(c) to increase the share of the school week devoted to examination subjects,
(d) to lengthen the school day and school year,
(e) to use specialist teachers to teach examination subjects,
(f) to hire teachers with a thorough background in the field,
(g) to reduce class size in examination subjects,
(h) to give teachers additional preparation time,
(i) to pay higher salaries,
(j) to spend more per pupil.

While senior secondary schools should be most affected, middle schools will also be influenced. If parents are able to choose which secondary school their child will attend and school funding is based on enrollment, the pressure on school administrators to provide a high-quality academic program will be particularly intense.

Assuming that administrators know how to allocate school resources to improve achievement, our 16th hypothesis is that:

H.16 — School resources and policies that regression analysis finds are effective should be more common in educational systems with external exams.

CHAPTER 4

TESTING FOR BACKWASH EFFECTS OF EXAMINATION SYSTEMS: INTERNATIONAL COMPARISONS

Do Examination Systems Affect Teacher Salaries, Qualifications and Spending Per Pupil?

Only a few industrialized countries (the U.S., Sweden, Spain, Portugal, and Belgium are the major examples) lack curriculum-based external exams at the end of secondary school (Kreeft, 1990, Madeus & Kellaghan, 1991; Eckstein & Noah, 1993; OECD, 1991). The theory presented in Chapters 1 and 3 predicts that these nations will

— set lower minimum qualifications for becoming a secondary school teacher,
— pay their secondary school teachers less (relative to other workers),
— allocate a smaller share of the education budget to teaching,
— have fewer teachers per pupil, and
— spend less overall on secondary education.

Table 4.1 presents data on salaries, qualifications and spending in 17 OECD countries that allows us to test these hypotheses. Since many countries fund pensions and medical insurance through mandated social security taxes, it is essential to include both voluntary and compulsory contributions for these purposes in the measurement of teacher compensation. Compensation of secondary teachers was calculated by multiplying their salary by the ratio of compensation to wages for manufacturing workers. This estimate of teacher compensation was then divided by average compensation of all workers (Nelson & O'Brien, 1993, pp.37, 74, 93).

Data on the relative wage of experienced lower and upper secondary teachers with 15 years of teaching experience are presented in the first two columns of Table 4.1. The countries with curriculum based external exams (CBEEs) pay experienced lower secondary teachers 58% more on average than the average worker. Experienced upper secondary teachers are paid 81% more. The four countries without a CBEE pay their experienced lower secondary teachers only 28% more and their upper secondary teachers only 50% more than the average worker. Thus, relative to others in their society, secondary school teachers in the CBEE nations are paid a statistically significant 21% more than teachers in the nations which lack curriculum-based external exams.

To some degree the salary advantage is compensation for longer teaching hours (see column 4, Table 4.1). The teachers in the CBEE nations spend on average about 10% more time providing instruction than teachers in no CBEE nations (Nelson & O'Brien, 1992).

As hypothesized, nations with curriculum-based external exams tend to set higher minimum standards for becoming a secondary school teacher. The schooling required to teach in a lower secondary school averaged 16.8 years in CBEE nations, significantly more than the 15.5 years in no CBEE nations. The more stringent entry qualifications are probably an important reason why teachers in CBEE nations are paid more.

Table 4.1.
Spending per Pupil and Teacher Pay and Qualifications

	Teacher wage (15 yrs experience)/Earnings All Wkrs[1]		Years of Education Req to Teach[2]		Yearly Hours Instruction[3] Upper S.S.	Secondary Pupils/ Teachers[4]	Spending Per Pupil/		Share of non teachers[6]
	Lower	Upper	Lower	Upper			Compensation Worker[5]	GDP Pop[5]	
No external exam									
Belgium	1.25	1.61	15	16	637	7.8	0.191	0.285	20%
Spain	1.40	1.72	15	17	536	16.6	0.107	0.218	
Sweden	1.14	1.32	16	16	593	13.0	0.243	0.364	
United States	1.33	1.33	16	16	825	15.9	0.201	0.283	56%
Mean — No Exam	1.28***	1.495**	15.5**	16.3*	648	13.3	0.1855	0.2875	38%
External exam									
Austria	2.01	2.48	15	16	819	9.4	0.270	0.356	
Australia	1.60	1.60				12.9			31%
Denmark	1.16	1.83			615	9.7	0.226	0.280	42%
Finland	1.61	1.67	18	18	556		0.195	0.333	46%
France	1.44	1.66	16	16	532	14.3	0.203	0.293	40%
Germany	1.82	1.98	19	20	810	16.2	0.243	0.305	18%
Ireland	1.49	1.49	17	17	801	17.1	0.123	0.217	
Italy	1.31	1.35	17	17	745	8.9	0.204	0.271	24%
Japan	1.71	1.73			696	16.6	0.182	0.202	23%
Netherlands	1.58	2.32	17	17	943	18.8	0.149	0.195	20%
Norway	1.33	1.65	15	16	586	8.3	0.311	0.351	
Switzerland	1.85	2.13			716				
United Kingdom	1.63	1.63	17	17	776	15.2	0.187	0.278	
Mean — Exam	1.58	1.808	16.8	17.1	716	13.4	0.2084	0.280	30.7%

[1]Compensation of secondary teachers was calculated by multiplying their salary by the ratio of compensation to wages for manufacturing workers. This estimate of teacher compensation was then divided by average compensation of all workers. The figure for French upper secondary teachers is a weighted average of salaries for Agrege (20%) and others (80%) (Nelson & O'Brien, 1993, pp. 73, 74, 90, 91).
[2]Minimum number of years of education required to be a lower or upper secondary teacher (OECD, 1995, p. 185).
[3]Mean number of hours teaching a class per day times the mean number of workdays for teachers (Nelson & O'Brien, 1993, Tables II.3 and II.4).
[4]The ratio of the number of full-time-equivalent pupils enrolled in public and private secondary schools to the number of full-time-equivalent secondary school teachers (OECD, 1995, p. 179).
[5]1992 Spending per secondary student divided by compensation per employee or per capita GDP (OECD, 1995, p. 90; Nelson & O'Brien, 1993, p. 91).
[6]Share of all staff employed in public or private primary, secondary and tertiary education that are not classroom teachers. The nonteaching staff includes administrators at all levels, teachers aides, guidance counselors, librarians, nurses, custodial staff, food service workers, bus drivers, and clerical workers. Figures for most nations are for all three levels of schooling (OECD, 1995, p. 176).

Contrary to our hypothesis, the ratios of secondary school students to full time equivalent secondary school teachers are not lower in nations with a CBEE system (see column 6, Table 4.1). Relative to per capita GDP, spending per secondary school student is also not higher in CBEE nations (see column 8, Table 4.1). Per capita GDP may not be the best deflator for education costs, however. Since labor compensation accounts for the bulk of education costs, the proper deflator for schooling expenditure is a wage index that reflects the cost of recruiting competent teachers. When spending per secondary school student is deflated by average compensation of all workers (see column 7, Table 4.1), the CBEE nations are estimated to spend about 12% more on their secondary schools than the no CBEE nations.

Our hypothesis that CBEE nations tend to focus school spending on teaching is also supported. Column 9 of Table 4.1 presents estimates based on population census data of the share of people working in all levels of education that are not teachers. Non teachers account for 56% of the employees of primary, secondary and tertiary education in the U.S. By contrast, in Germany, Italy, Japan and the Netherlands, non teachers account for less than a quarter of education employees (see Column 5 of Table 4.1). The average for the eight CBEE countries for which data are available was 30.7%; substantially lower than the 38% average for Belgium and the U.S., the two no CBEE countries for which data were available.

Nelson's (1990) examination of this issue also concludes that teacher compensation is a smaller share (between 45.5 and 53.5%) of current expenditures in the U.S. than in most other OECD countries. The mean for OECD countries reporting this statistic was 68% (OECD, 1993, p.88).

Impacts on Student Achievement in Math, Science and Geography: An Analysis of Data from the International Assessment of Educational Progress

Three of the countries that lack curriculum-based external examinations — Spain, Portugal and the United States — participated in the 1991 International Assessment of Educational Progress (IAEP). Consequently, a comparison of student achievement in these three countries to that in the rest of the IAEP sample provides an initial test of the hypothesis that CBEEs raise student achievement. The power of this test, however, is very low. Only 20 nations participated in the IAEP science and mathematics study. Low levels of industrialization necessitate the exclusion of four nations — Brazil, China, Jordan and Mozambique. Canada participated as nine different provinces, five of which were stratified into separate English speaking and French speaking school systems. The Canadian data on mathematics and science achievement were set aside for separate analysis in Chapter 6. The 15 countries included in the analysis are: England, France, Hungary, Ireland, Israel, Emilia Romagna/Northern Italy, Korea, Portugal, Scotland, Slovenia, Soviet Union, Spain, Switzerland, Taiwan and the United States.

In the IAEP, schools were first sampled, then students within schools. Sampling frames generally excluded separate schools for special education students and often very small schools as well. Israel assessed only its Hebrew speaking schools, the Soviet Union assessed Russian language schools in 14 of the nation's 15 republics. Switzerland's assessed 15 of 26 cantons. A school's likelihood of selection was roughly in proportion to its estimated number of 13 year olds. In most countries school non response rates were extremely low. They were zero in Hungary, Slovenia, Korea and Taiwan and 3% in Israel and the Soviet Union. The countries with high non response rates were Switzerland (17%), Emilia Romagna (18%), Scotland (19%), U.S.A. (21%) and England (48%). When sampled schools declined to participate an alternative was selected from the same stratum (IAEP, 1992c).

Random samples of 30–34 thirteen year olds were selected from each school. Half were assigned to the mathematics assessment and half assigned to the science assessment. Students also completed a brief questionnaire that asked about books in the home, number of siblings, language usually spoken at home and home availability of calculators and computers.

The following procedure was implemented to standardize for differences in socioeconomic background across countries. Cross-section regressions were run within each nation predicting mathematics and science test scores adjusted for guessing as a function of gender, number of siblings — 2, age — 13.5, and dummies for language of instruction different from language used at home, 10 or fewer books at home, 11–25 books at home, and more than 100 books at home. The mathematics achievement regression also had dummy variables for "have a calculator," and "ever use a computer for school work or homework." The resulting regressions were then used to predict adjusted percent correct for three groups of male and female students: those from families with 10 or fewer books, those with 25–100 books and those with more than 100 books. The predictions were made for students who were 13.5 years old, had 2 siblings and spoke the national language at home. The mathematics predictions assumed that the student had a calculator but not a computer. A weighted average of predicted percent corrects was then

calculated with male and female students with 10 or fewer books each given a weight of 0.1 and the 25–100 books and over 100 books predictions for each gender each getting a weight of 0.2.

This average adjusted percent correct was then regressed on per capita gross domestic product (GDP deflated by a purchasing power parity price index), a dummy for Asian nation (Korea or Taiwan) and a dummy for curriculum-based external examinations. The results are presented in Table 4.2.

Mathematics

Students in Korea and Taiwan score significantly higher than students in Europe and North America: 17 points higher in mathematics and 12–13 points higher in science. The U.S. standard deviation was 26.8 percentage points in mathematics and 19.9 in science, the Korea/Taiwan effect is more than one half a U.S. standard deviation. Coefficients on per capita GDP are positive but not statistically significant. The low-income countries in this sample are Hungary and the Soviet Union and their students do remarkably well on the IAEP tests.

Of primary interest are the estimated impacts of curriculum-based external examination systems. For mathematics the effect is highly significant and quite large (about one-half of a U.S. standard deviation or nearly two U.S. grade level equivalents). The third row in Table 4.2 presents

Table 4.2.
Impact of Curriculum-Based Examinations at the End of Secondary School on Mathematics, Science and Geography Achievement
Age 13 — IAEP

	Curriculum-based exam	Asian nation	1990 GDP/Pop (PPP)	USA	Spain	Portugal	Canada No Exam	Canada Exam	R^2 Rmse	Number of observations
Mathematics										
Adjusted Avg.	13.8***	14.8***							0.640	15
	(3.1)	(3.4)							6.1	
Adjusted Avg.	14.2***	16.3***	0.45						0.643	15
	(3.5)	(3.3)	(1.1)						6.1	
Adjusted Avg.		16.7***	0.56	-17.1*	-10.9	-14.8*			0.583	15
		(3.1)	(1.0)	(2.2)	(1.6)	(2.1)			6.6	
Raw Avg.	15.6***	14.1***	0.46						0.649	15
	(4.0)	(2.9)	(1.1)						5.9	
Science										
Adjusted Avg.	1.9	11.0***							0.584	15
	(0.9)	(4.3)							3.3	
Adjusted Avg.	2.2	12.1***	0.34						0.629	15
	(1.1)	(4.8)	(1.6)						3.1	
Adjusted Avg.		11.9***	0.27	-1.2	-0.3	-4.9			0.593	15
		(4.4)	(1.0)	(0.3)	(0.1)	(1.4)			3.3	
Raw Avg.	4.7	9.4**	0.13						0.439	15
	(1.8)	(2.9)	(0.4)						4.0	
Geography										
Adjusted Avg.	3.5**	-2.4							0.146	21
	(2.3)	(0.7)							3.3	
Adjusted Avg.	2.9	-3.3	-0.15						0.127	21
	(1.7)	(0.9)	(0.8)						3.4	
Adjusted Avg.		-2.3		-0.8	-2.3		-3.8*	0.2	0.023	21
		(0.6)		(0.2)	(0.6)		(1.9)	(0.1)	3.6	
							F=3.9	(prob= 0.068)		

***P <0.01 on a two-tail test.
**P <0.05 on a two-tail test.
*P <0.10 on a two-tail test.

estimates of how each no CBEE country deviates from the prediction of the model. Portugal, Spain and the U.S. underperform by similar amounts.

Science

For science the impact of curriculum-based external examinations is positive (about 12% of a U.S. standard deviation) but not statistically significant. When each no CBEE country is allowed to have its own separate effect, Portugal is about 25% of a U.S. standard deviation below the level predicted by the regression. The coefficients on the Spanish and U.S. dummies are negative but tiny.

Table 4.2 also presents regressions predicting mean achievement levels unadjusted for demographic differences between countries. Surprisingly, this does not increase the estimated effect of per capita GDP on achievement. Curriculum-based external exams have slightly larger effects on achievement in these models. Impacts on science achievement just miss significance at the 5% level on a one-tail test.

Geography

Nine of the countries in the IAEP study assessed geography as well as mathematics and science. The countries participating in the geography assessment were Canada, Hungary, Ireland, Korea, Scotland, Slovenia, Soviet Union, Spain and the United States. Canada collected sufficient data to allow valid comparisons between provinces and between the Anglophone and Francophone school systems of the five provinces with dual education systems. Some of these provinces have CBEEs and others do not, so including the Canadian provincial data in the study substantially increases the power of our tests for the effects of exams (IAEP, 1992d).

Regressions were estimated predicting country means using the same specifications as above. Per capita GDP had the wrong (a negative) sign, so preferred specifications are those which do not include this variable. In the preferred model (row 9 of Table 4.2), curriculum-based exams have a significant effect on geography achievement. The effect appears to be roughly 20% of a U.S. standard deviation. Row 11 presents a regression where each of the no CBEE countries has its own coefficient. The U.S. does not lie below the regression line. The Canadian provinces without examination systems do significantly worse than the provinces that have examinations systems (F for Hyp = 3.9).

These results are consistent with the causal hypotheses presented above. Causation is not proved, however, because other explanations for the U.S., Spanish and Portuguese lag can no doubt be proposed. Other sources of variation in curriculum-based exams need to be analysed. Best of all would be studies which hold national culture constant.

Effects of Sweden's Elimination of Exit Examination

Another approach is to study the effects of changes in examination systems. There are, however, three problems with this approach. The first problem is that changes in examination systems will influence student achievement outcomes only with a long lag. An 18 year old's

skills in mathematics are the product of 12 years of education, not just what happened during the previous year. But even more important is the conservative nature of schools as institutions. School cultures are very slow to change. Teachers who have adapted to a system of external exams by setting high standards for their students will not quickly lower their standards when they are given more authority over final senior year grades. Parental and public attitudes about learning and the proper goals and priorities for schools also shift at a glacial pace. Consequently, it will probably take more than a decade for most of the effects of changes in examination regime to be felt. Over that time frame schools and societies change in so many other ways; it will be hard to distinguish the effects of changes in examination regime from the effects of other phenomena.

The second problem is the marginal character of most of the changes that have been made. Many countries have tinkered with the mechanism for deciding who is admitted to different kinds of post-secondary education or have lowered the weight of external exam when assigning course grades. As long as the external exams count in the final grades and college admissions decisions, the weight attached to them may not be all that critical. The passing standard of the exams affects their incentive impacts, but the passing standard is determined endogenously with average performance levels, so one cannot treat the passing standard as an exogenous determinant of performance.

The third problem is the limited number of countries for which it is possible to track performance levels over time. Sweden appears to be the only country for which such data are available which has made a non marginal change in its examination system.

The Case of Sweden

Sweden eliminated its high-stakes curriculum-based external examinations at the end of secondary school during the 1970s (Eckstein & Noah, 1993). Certification of achievement was now to be based on continuous assessment by teachers. Fearing grade inflation, however, the National Board of Education developed the multiple choice and short answer *centrala prov* subject examinations "to help teachers grade students properly .... Teachers have to use them and are not allowed to deviate more than 0.2 grade points from the standardized test class means in their final evaluation (Kreeft, 1990, p. 15)." While the grades awarded by teachers were supposed to follow a normal distribution centered on the class mean on the *centrala prov*, there is controversy about whether these mandates were being universally followed. The Swedish students I have talked to did not perceive the *centrala prov* exams as carrying high stakes for themselves, so it fails requirement 1 for being a CBEE.

Sweden not only ended its exit exams, it established new rules for selecting applicants for university. Swedish upper secondary students pursue specialized lines of study beginning in 10th grade. Universities had traditionally admitted recent graduates of academic lines of study in upper secondary school with three year durations. Reforms introduced in the 1970s changed the rules governing competition for university places to favor those who had worked a few years after graduating from secondary school (Rehn, 1980, p. 150). This induced many secondary school graduates to postpone entry into higher education in order to accumulate enough work points to gain admission to preferred educational programs. Entrants to the study of medicine, for example, had an average age of 30 in 1977 and 1978 (Rehn, 1980, p. 154).

In addition, universities were not allowed to give preference to secondary school graduates who had pursued the more difficult longer duration academic lines of study in upper secondary

school. If one-third of all applicants to university had pursued a two-year vocational line of study, the universities and colleges were required to admit one-third of their entering class from this group. This meant that a given student's chances of being admitted to popular majors in university were generally higher if the student pursued a less demanding shorter duration vocational line of study. As a result, in Fall 1978, the toughest secondary school major, natural science, was very easy to get into. Ninety-three percent of those who selected it as a first choice were admitted. For those who wanted to enter a two-year vocational line in secondary school, only 46% were admitted to their first-choice program.

These changes in secondary school examinations and university admissions policies appear to have caused a decline in the number of Swedish upper secondary students taking rigorous courses in mathematics and science during the 1970s.

Mathematics at graduation

Eight countries participated in both the First International Mathematics Study in 1964 and its replication in 1982 (see Fig. 4.1). **The proportion of Swedish 18 year olds taking college-prep mathematics fell from 16% to 12%. This more selected group of students scored only slightly higher on the anchor items that appeared in both assessments.** Finns, by

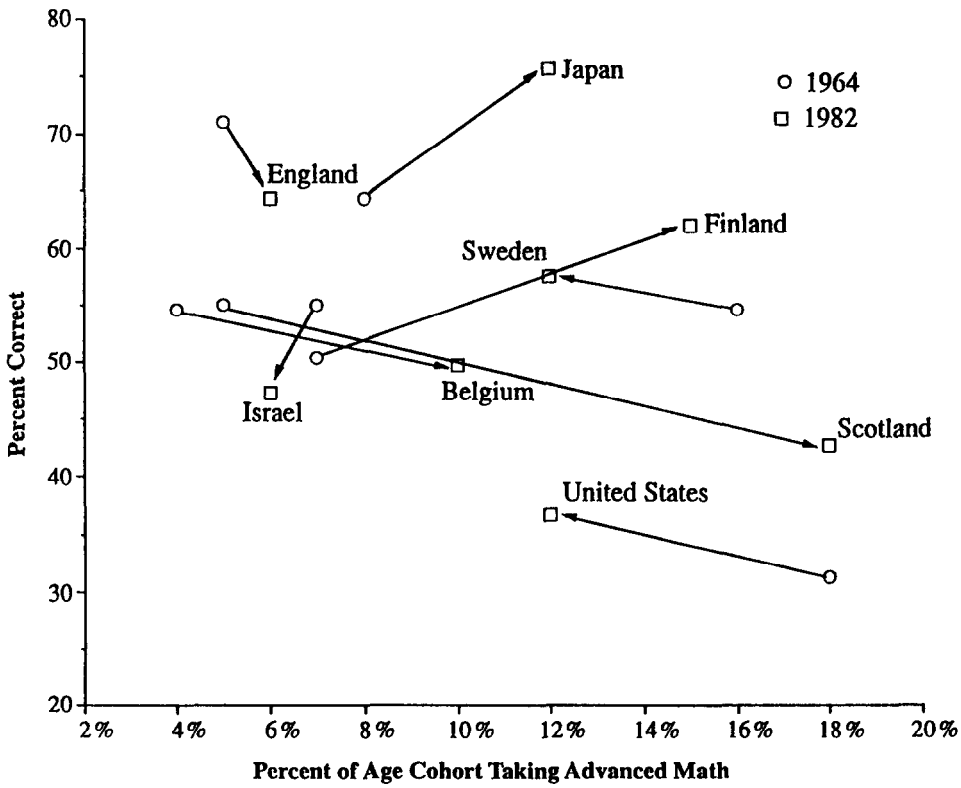

Figure 4.1. Mathematics at the end of high school.

contrast, simultaneously increased the proportion of the age cohort taking college prep mathematics from 7 to 15% and significantly improved their mean scores (Husen, 1967; Robitaille & Garden, 1989).

Science at graduation

A similar analysis was conducted of changes in Swedish performance on IEA science examinations given at the end of secondary schooling. The data necessary to measure changes in absolute levels of achievement between the First and the Second International Science Studies are not available. Comparisons of achievement relative to that of other countries are possible, however, and are presented in Fig. 4.2. The zero point on the scale for each year is the average performance *for that year* of the English students preparing for A level exams. England was selected for this standardization role because there was no change in the proportion of its 19 year old age cohort who were in the sample frame for the IEA science study. The standard deviation used as the metric measures individual variance for pooled data on 14 year olds from 11 (14 in 1983) industrialized countries. The vectors describe how a nation's scores and participation rates changed relative to England. The 1983 study reports separate means for non specialists (e.g., majors in economics, languages) and for students specializing in science. These two figures were averaged using the shares of the age cohort who pursue these two courses of study as weights (Postlethwaite & Wiley, 1992, p. 6, 74).

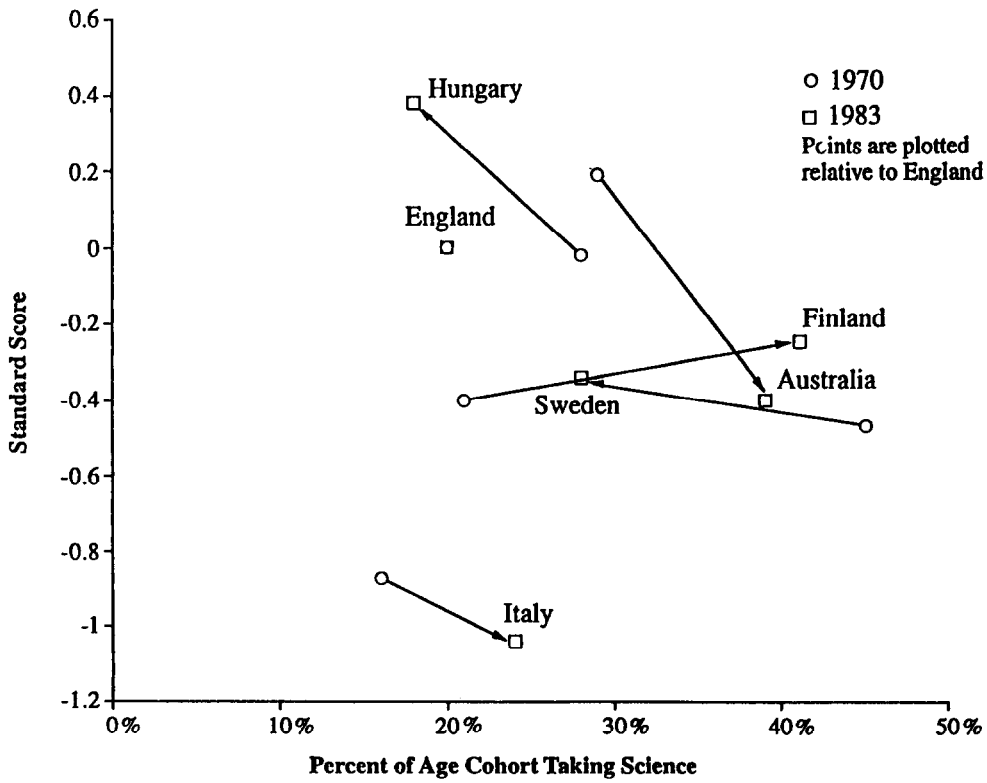

Figure 4.2. Science at the end of high school.

The proportion of Swedish 18 year olds in academic lines of study that were given the IEA science exam fell from 45% to 28% between 1970 and 1983. This much smaller and surely more able group of Swedish students only slightly improved their position relative to England (see Fig. 4.2). Finland, by contrast, increased the proportion of the age cohort assessed in science from 21 to 41% and simultaneously improved its score relative to England. Other countries — Australia, Italy — had declines in relative test scores, but they were associated with large increases in the proportion of 18 year olds in the academic lines of study that were tested. Hungary was the only other country to reduce the share of 18 year olds that participated in the IEA science exam and it experienced a substantial increase in relative test scores (Comber & Keeves, 1973, p. 168; Postlethwaite & Wiley, 1994, p. 74).

Science at age 14

The decline in Sweden's relative performance appears to be an upper secondary school phenomenon. The relative performance of Swedish 14 year olds did not decline between 1970 and 1983/84. Of the eight countries that participated in both the First and Second International Studies of Science achievement of 14 year olds, four lost ground on Sweden and three gained ground. Japan and Hungary originally way ahead remained ahead, but their lead diminished. England and the United States were roughly equal to Sweden in 1971, but fell back in the succeeding decade. Finland and Italy were behind in 1971 but had essentially caught up by 1982. Dutch 14 year olds were way behind in 1971, but had surpassed Sweden by 1982.

Mathematics at age 14

Mathematics achievement of Swedish 14 year olds was low in 1964 and remained low in 1982. Changes in mathematics achievement can be calculated for nine other countries. Three improved mathematics achievement levels: Finland, France and the Netherlands. In Belgium and Israel large declines in test scores were associated with a significant increase in the proportion of 14 year olds in school between 1964 and 1982. The other four countries with declining test scores had close to 100% of 14 year olds in school in 1964. The test score declines were modest for Japan and the U.S. but substantial for England and Scotland. Mathematics achievement declined 8.4 points for England and 4.4 points for Scotland even though enrollment rates were stable at high levels.

The lack of backwash effects on Swedish 14 year olds is probably due to the fact that selection for preferred specializations in upper secondary school is only a year or so away. Which specialization is preferred has changed over time, but the fact of competition has not. Certain lines of study are strictly rationed. For example, in 1978 only 22% of those giving nursing as their first-choice specialization were admitted to the line. Admission to rationed lines is based on grades in lower secondary school (Rehn, 1980), so external incentives to study in lower secondary school were maintained even while they were being weakened in upper secondary school.

CHAPTER 5

NEW YORK STATE vs. THE REST OF THE UNITED STATES

Introduction

Probably the best way to test these hypotheses is to limit the study to just one country and compare jurisdictions with different kinds of examination systems. The jurisdictions must be large, however, for otherwise colleges and employers are not likely to use grades on the curriculum-based exams in their selection decisions, so the rewards for doing well may be quite limited.

New York State is reasonably large and has a curriculum-based Regents Examination system. It began in 1865 as a system of high school entrance examinations. "The plan of uniform and impartial entrance examinations was immediately successful, and there soon arose a strong demand for similar safeguards and standards for high school graduation and college admission" (The University of the State of New York, 1965, p. 1). The first high school examination exams were administered in June 1878. They have been in continuous evolution ever since.

The examinations are taken throughout one's high school career. A student taking a full schedule of college preparatory Regents courses would typically take Regents exams in mathematics and earth science at the end of 9th grade; mathematics, biology and global studies exams at the end of 10th grade; mathematics, chemistry, English and foreign language exams at the end of 11th grade and mathematics and physics exams at the end of 12th grade.

The system now reaches over half of the state's high school students. In 1994, about 56% of 9th graders took the Mathematics Course 1 exam and, of these, 24% failed. Similar proportions of 10th and 11th graders took the global studies, biology and English exams. Failure rates were 20% in global studies, 18% in biology and 13% in English. The great bulk of those not taking Regents exams are in courses that are considerably less challenging than Regents level courses. The fact that nearly half of New York students are avoiding Regents courses because they perceive them to be too much work or too difficult suggests that the standard of the exam is about as high as is feasible considering current average achievement levels in the state. A system of minimum competency tests in specific subjects sets a minimum standard for those not taking Regents courses. Of the 50 states, New York is the only one with such a system.

Does New York State Invest More in K-12 Education?

The theory predicts that New York State will pay its teachers more, employ more qualified teachers, spend more per student, have smaller classes and focus that spending on instruction. Table 5.1 presents data on education spending patterns from New York and 10 other large and nearby states. Stars indicate whether New York is significantly higher than the other 49 states. Indeed, New York pays significantly higher salaries than other states. It pays better than wealthier

Table 5.1.
Spending and Academic Outcomes New York vs. Other States

	New York	U.S . Avg.	California	Illinois	Ohio	Michigan	Wisconsin	Pennsylvania	New Jersey	Connecticut	Massachusetts	Maryland
Effort — Public												
K-12 Teacher Salary-92[1]	45,400**	35,900	43,300	37,800	35,400	39,900	34,900	39,400	43,600	48,900	44,500	40,400
% Teacher Sal/Faculty Sal[2]	91.3*	81.6	79.7	84.4	80.2	94.9	82.8	85.2	84.8	92.0	86.9	90.2
% Masters Deg.[3]	67.5**	47.2	44.5	47.7	40.1	59.6	36.6	50.4	40.5	75.0	51.3	50.8
90 Pupil/teacher[3]	14.7	17.2	22.8	16.7	17.2	19.8	16.5	16.6	13.6	13.6	15.4	16.8
% $ Instruction[4]	66.0**	60.4	57.8	59.1	56.6	56.7	62.4	61.4	59.1	63.5	58.9	59.4
% K-12 Spending/per cap GSP[5]	22.1	19.6	16.2	16.3	18.7	21.4	20.7	22.1	21.1	20.0	18.8	20.6
% K-12 Spending/Higher Ed.[6]	82.7***	52.0	46.4	64.3	53.7	56.3	54.2	54.0	76.1	78.2	75.5	65.7
Context												
89 Med Fam Inc[7]	33,000	30,100	36,800	32,300	28,700	31,000	29,400	29,100	40,900	41,700	37,000	39,400
% Poverty Fam[7]	10.0	10.0	9.3	9.0	9.7	10.2	7.6	8.2	5.6	5.0	6.7	6.0
% Black[7]	15.9	12.1	7.4	14.8	10.6	13.9	5.0	9.2	13.4	8.3	5.0	24.9
% Hispanic[7]	12.3	9.0	25.8	7.9	1.3	2.2	1.9	2.0	9.6	6.5	4.8	2.6
% Foreign Born[7]	15.9	7.9	21.7	8.3	2.4	3.8	2.5	3.1	12.5	8.5	9.5	6.6
Pop 25+ Literacy[8]												
% Prose Level 3+	46	52		52	55			54	53			
% Doc Level 3+	43	49		49	51			49	49			
% Quan Level 3+	46	52		54	56			54	54			
Student Outcomes												
90–91 Students Taking[8]												
% Algebra I	83	77	89	77	86		95+	88		81		94
% Algebra II	45	49	42	39	50		58	60		59		51
% Calculus	13	6.6	9	9	10		21	19		14		13
% Biology	95+	92	89	78	93		95+	92		95+		95+
% Chemistry	58	50	33	40	53		59	59		63		61
% Physics	28	22	15	20	22		27	31		37		27
% Take AP Eng, Sci, Math & Hist[8]	14.5**	8.5	11.1	8.6	5.6	6.9	5.2	7.2	11.9	10.9	11.6	12.7
% Pass AP Eng, Sci, Math & Hist[8]	9.4**	5.4	7.1	6.1	3.7	4.5	3.3	4.7	8.2	8.0	8.2	8.8
% Pass AP F.Lng[8]	1.0**	0.6	1.7	0.5	0.2	0.3	0.2	0.3	0.8	1.1	1.0	0.8

[1]Average annual salary of instructional staff 1991–92, (includes school administrators) (NCES, *Digest*, 1993, Table 76).
[2]Ratio of 1990–91 salary of K-12 instructional staff to salary of full-time instructional faculty on 9 month contracts in all higher education (NCES, *Digest*, 1993, Table 76, p. 220).
[3]NCES, 1992, Tables 62 and 64.
[4]Instruction expenditures divided by sum of instruction, support and noninstruction (NCES, *Digest*, 1993, Table 154).
[5]Ratio of public K-12 spending per pupil to gross state product per capita (NCES, *States and Nations*, 1993, Table 15b).
[6]Ratio of public K-12 spending per pupil in average daily attendance to education and general expenditures at public institutions of higher education per FTE student (NCES, *Digest*, 1993, Tables 157, 189, 333).
[7]Data are from the 1990 census (Slater & Hall, 1995, pp. 3, 6).
[8]National Education Goals Panel, 1993, Volume 2. AP data are the ratio (in % metric) of the number of AP exams taken/passed to the number of 11th and 12th graders in state.
*** $P < 0.01$ on a one-tail test, ** $P < 0.05$ on a one-tail test, *$P < 0.10$ on a one-tail test.

states such as Massachusetts, New Jersey, New Hampshire and Maryland (see row 1). Only Alaska and Connecticut pay more. It is not just a high cost of living that produces this effect. New York's ratio of K-12 teacher salaries to college faculty salaries is also significantly (at the 10% level) above the national average (see row 2). New York teachers are also more likely to have master's degrees than the teachers of any state except Connecticut and Indiana. New York ranks number 7 in teacher–pupil ratios (row 4) and number 7 in the ratio of spending per pupil to gross state product per capita (row 6).

New York is also significantly above average in the share of K-12 spending that it allocates to instruction (row 5). Only Delaware exceeds the New York level. Much of the extra spending is on teachers' aides and other specialized instructional staff. In New York, teachers accounted for 51% of school staff in 1990. The national average was 53.3% (NCES, 1992, Table 79).

Clearly, New York invests a great deal in its K-12 education system. If the cause of the high spending were a strong general commitment to education or legislative profligacy, we would expect spending to be high on both K-12 and higher education. This is not the case. New York is number 1 in the ratio of K-12 spending per pupil to higher education spending per college student (row 7). Compared to other states, New York gives K-12 education greater priority than higher education.

New York's heavy investment in K-12 education is not a result of a strong tax base. In 1989 New York's median family income was only 10% above the national average (row 8). Its poverty rate was equal to the national average (row 9). The neighboring states that also have high teacher salaries and high spending per pupil, such as Connecticut, New Jersey and Massachusetts, are much wealthier. Connecticut and New Jersey's median family income was 24–26% above New York's median in 1989.

Do New York Schools Add More Value than the Schools of Other States?

New York's student population is more disadvantaged, more heavily minority and more likely to be recent immigrants than the students of most other northern states. Among northern states, only Maryland and Delaware have a larger share of African-American citizens (row 10). Nationwide, only California has a higher share of its population foreign born and only California, Texas, Arizona, New Mexico and Colorado have larger Hispanic population shares (rows 11 and 12). Literacy levels among adults are substantially below the national average (rows 13–15).

Consequently, when one compares student achievement levels, family background must be taken into account. Considering the disadvantaged character of its students, New York students do remarkably well. The proportion of students taking algebra, calculus, chemistry and physics is generally above national averages but below Pennsylvania, Connecticut, Maryland and Wisconsin and a number of other states (rows 16–21). With respect to participation in Advanced Placement classes, New York is significantly higher than the rest of the nation. A larger proportion of its 11th and 12th graders are taking and passing (9.4%) AP exams in English, science, math or history than any other state except Utah (rows 22–24).

The Impact of Curriculum-Based Exams on SAT Scores

Our theory predicts that, when demographic characteristics are held constant, New York students should do better on secondary school exit exams than students from other states. As

predicted, New York students do out perform students in other states. Graham and Husted (1993) discovered this fact when they examined the determinants of mean SAT test scores in the 37 states with reasonably large test-taking populations. Controlling for the proportion of high school seniors taking the SAT test and the race, gender, parental income and parental education of test takers, they found that New York State had the highest adjusted mean Scholastic Aptitude Test score. They did not, however, test the statistical significance of the New York State effect and used an unusual log–log specification.

Are their findings robust to changes in specification? How large is New York's advantage? Is the differential statistically significant? Table 5.2 presents the results of a regression predicting 1991 mean Math SAT-I, Verbal SAT-I and Total SAT-I test scores for the 37 states for which data are available. With the exception of the dummy variable for New York State, all right-hand side variables are proportions — generally the share of the test-taking population with the characteristic described. Clearly, New Yorkers do significantly better on both the math and verbal sections of the SAT-I than students of the same race and social background living in other states. When this model is estimated without the NYS dummy variable, New York has the largest positive residual in the sample. The next largest (Wisconsin's) positive residual is 87% of New York's residual. Illinois and Nevada have positive residuals that are about 58% of New York's value. Arizona, California, Colorado, Florida, New Mexico, Ohio, Rhode Island, Texas and Washington have negative residuals greater than 10 points. Many of these states have large populations of Hispanics and recent immigrants, a trait which was not controlled for in the analysis. This makes New York's achievement all the more remarkable when one considers that Hispanics are 12.3% of the state's population and that only California has a higher share of its population foreign born.

Adding the logarithm of the teacher–pupil ratio and spending per pupil to the model reduces the NYS coefficient by 25% (from 46 in row 3 to 35 in row 4). It remains significantly greater than zero, however. This suggests that one of the reasons why New York students do better on the SAT than students in other states is the heavy investments in K-12 education made by the taxpayers of the state. Most of the effects of the Regents exams, however, operate through other mechanisms such as higher standards, studying harder, etc.

For individuals the summed SAT-V + SAT-M has a standard deviation of approximately 200 points. Consequently, the differential between New York State's SAT mean and the prediction

Table 5.2.
Determinants of Mean SAT Scores for States

	NYS	Partic rate	Parents AA-BA+	Private school	Prop. Black	Large school	3+ Math courses	3+ Eng. courses	InTeach/ stud	InExp/ stud	R^2 RMSE
Math	25**	-45***	173***	46**	-83***	-8	42	-61			0.916
SAT-1	(2.7)	(3.1)	(5.3)	(2.2)	(3.5)	(0.6)	(1.1)	(0.9)			8.3
Verbal	21**	-23	197***	14	-52**	-36**	43	26			0.897
SAT-I	(2.1)	(1.5)	(5.9)	(0.6)	(2.1)	(2.5)	(1.1)	(0.4)			8.6
Total	46**	-68**	370***	60	-135***	-44*	85	-36			0.926
SAT-I	(2.7)	(2.6)	(6.4)	(1.6)	(3.2)	(1.8)	(1.3)	(0.3)			14.8
Total	35*	-88***	367***	69*	-113***	-36	45	-45	48*	13	0.933
SAT-I	(2.0)	(3.3)	(6.6)	(1.9)	(2.6)	(1.5)	(0.7)	(0.4)	(1.7)	(0.8)	14.2
											Total SAT
Mean	0.027	0.414	0.581	0.207	0.078	0.129	0.617	0.797	-2.822	1.648	925
SD	0.164	0.240	0.097	0.082	0.064	0.113	0.067	0.038	0.113	0.215	55

***P <0.01 on a two-tail test.
**P<0.05 on a two-tail test.
*P <0.10 on a two-tail test.

for New York based on outcomes in the other 36 states is about 23% of a standard deviation or about three-quarters of a grade level equivalent.

This occurred despite the fact that Regents exam grades generally account for less than half of final grade in the course and influence only the type of diploma received. Employers ignore exam results when they make hiring decisions. During the 1980s, scholarships sponsored by the Regents were based on aptitude test scores, not Regents exam results. A passing score on a Regents exam is not necessary for admission to community colleges or out of state colleges. Students are aware that they can avoid Regents courses and still go to college. Indeed some perceive an advantage to avoiding them:

> My counselor wanted me to take Regents history and I did for a while. But it was pretty hard and the teacher moved fast. I switched to the other history and I'm getting better grades. So my average will be better for college. Unless you are going to a college in the state, it doesn't really matter whether you get a Regent's diploma. (Ward, 1994).

Indeed, the small payoff to taking Regents exams may be one of the reasons why nearly half of students take local rather than Regents courses or take the Regents course but skip the exam.

CHAPTER 6

MATHEMATICS AND SCIENCE ACHIEVEMENT IN CANADA: COMPARING PROVINCES

Introduction

Probably the best place to test hypotheses about the impact of curriculum-based external examinations is Canada. Some Canadian provinces have curriculum-based exams, others do not. At the time the data used in this study were collected, 1990–91, Alberta, British Columbia, Newfoundland and Quebec had curriculum-based criterion-referenced examinations in English, French, mathematics, biology, chemistry, physics and many other subjects. These exams accounted for 50% of the final grade in Alberta, Newfoundland and Quebec and 40% in British Columbia. New Brunswick had such exams in language arts and mathematics but not in science or other subjects. Local school districts decided how much weight the New Brunswick Provincial Achievement Exam was to be given in assigning final grades.

The other provinces had no curriculum-based provincial examinations in 1990–91. Ontario eliminated them in 1967, Manitoba in 1970 and Nova Scotia in 1972. Nova Scotia substituted multiple-choice norm-referenced achievement tests in reading, language usage, proof-reading, mathematics, science and social studies which do not influence student grades. In Ontario, some local school districts have district level exams for core subjects, but most do not. In any case, one would not expect local district subject exams to have as powerful incentive effects as provincial or national exams.

Figure 6.1 plots provincial mean mathematics achievement of 16 year olds from the 1993 School Achievement Indicators Program (SAIP) against the mean literacy level of the parents of each province. The provinces with CBEEs — *Quebec, Alberta, British Columbia and Newfoundland* — are represented by squares and circles. Diamonds and triangles indicate the provinces without a CBEE — Ontario, Nova Scotia and Manitoba. Most CBEE provinces had higher scores on the SAIP assessments than the provinces without CBEEs with the same level of parental literacy (Council of Ministers of Education, Canada, 1993; Jones, 1993).

Three of the provinces that lacked CBEEs in 1991 have now reintroduced them (GAO, 1993). Manitoba introduced its 12th grade examination in the winter of 1991 about the time the IAEP exam was being administered to 8th graders in the province. The new examination system was announced in June 1990, only 7 months earlier. The teaching of 12th graders was probably affected during the 1990–91 school year, but it is unlikely that any incentive effects had percolated back to the 8th grade in so short a time. The system introduced in Manitoba in 1991 rotated the subject assessed on a five or six year cycle. Starting in 1996, Manitoba will assess math and language arts every year and require the exams to count for 30% of the student's final grade.

New Brunswick had a no stakes examination in mathematics and language arts in 1991.

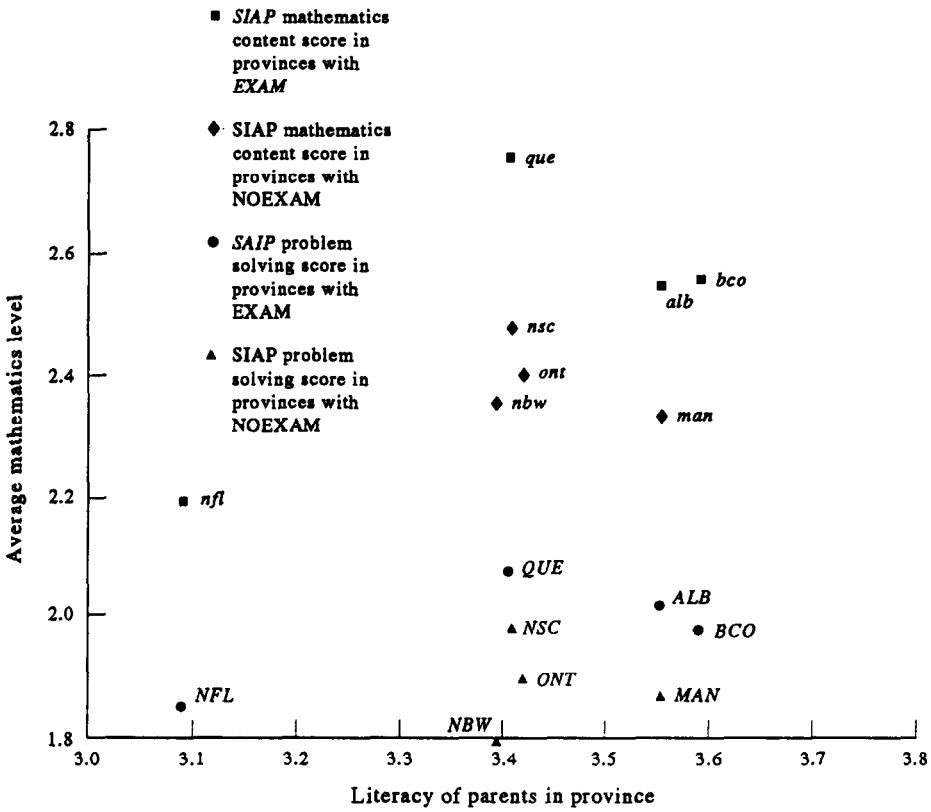

Figure 6.1. Mathematics achievement of 16 yr olds by curriculum-based exam system in province.

External exam grades were not separately reported on student transcripts and local districts were not required to average exam grades into the student's final grade. The result was the exams had only limited "credibility" and many students chose not to take them. Starting in 1995, however, the province requires that exam results account for 30% of course grades. Exam results will also soon be included on the transcript. In the past district and school averages of exam results were treated as confidential information made available only to the local superintendent. The provincial legislature is likely to change this practice in the near future, however. New Brunswick is also planning to extend its end of high school examinations to science. New Brunswick, Nova Scotia and Newfoundland are cooperating in the development of examinations for Biology, Chemistry and Physics. Alberta's examination system was instituted in 1984, so it was 7 years old when the IAEP data were collected.

In terms of governance and finance, the Canadian system of elementary and secondary education is more similar to America's system than almost any other nation. Comprehensive schools predominate. As in the U.S., education is a provincial/state responsibility. Localities administer schools and use the property tax to raise their share of the funding. In 1980 localities accounted for 43% of the funding in the United States and 28.5% in Canada. Funding levels vary less within Canadian provinces than within American states. The average within province coefficient of variation is 0.09 for Canada and 0.17 for the United States (McDonald, 1994; National

Center for Education Statistics, 1992). In some provinces negotiations over teacher salaries occur at the provincial level.

The analysis proceeds as follows. The data are described in the next section, and an analysis of the determinants of math and science test scores is presented in the following section. The impact of provincial exam systems on the behavior of parents, students, teachers and school administrators is presented in the following section.

The IAEP Data

The hypotheses outlined in Chapter 3 will be tested on data on the mathematics and science competence of 42,241 Canadian and American 13 year olds from the International Assessment of Educational Progress (IAEP). When the Educational Testing Service canvassed countries about participating in the 1991 International Assessment of Educational Progress, Canada decided to collect sufficient data to allow valid comparisons between provinces and between the Anglophone and Francophone school systems of the five provinces with dual systems. The Yukon, the Northwest Territories and Prince Edward Island did not participate in the study. Stratified random samples of 105–128 secondary schools were selected from the French speaking school systems of Ontario and Quebec and from the English speaking school systems of Alberta, British Columbia, Manitoba, Saskatchewan, Ontario, Quebec, New Brunswick, Nova Scotia and Newfoundland. The United States sample contained 106 schools. A school's likelihood of selection was roughly in proportion to its estimated number of 13 year olds. All French speaking schools in New Brunswick, Saskatchewan and Manitoba were invited to participate.

In the United States 21% of the schools invited to participate declined. In Canada the highest school refusal rates were for the English speaking schools in Quebec (15%), the English speaking schools in Saskatchewan (12%) and the French speaking schools in New Brunswick (12%). In the rest of the provinces refusal rates were below 7% and in many provinces all invited schools participated. When sampled schools declined to participate, an alternative was selected from the same stratum (IAEP, 1992c).

Random samples of 30–34 thirteen year olds were selected from each school. Half were assigned to the mathematics assessment and half assigned to the science assessment. Students also completed a brief questionnaire that asked about books in the home, number of siblings, language usually spoken at home, hours watching TV, hours doing homework, pleasure reading, watching science programs on TV, home availability of mathematics and science resources, and attitudes towards math and science. The principals of participating schools also completed questionnaires describing school policies, school resources and the qualifications of 8th grade mathematics and science teachers.

Impacts of Curriculum-Based Examinations on Math and Science Achievement

In this first stage of the analysis the dependent variable is the percent correct with adjustments for guessing. It is defined as the number of correct answers minus 0.25 times the number of answered questions all divided by 0.75 times the number of items on the test. Adjusted for guessing, the students examined here got an average of 47.2% in math (SD=24 points) and 57.3% in science (SD=20.2 points). The independent variable of greatest interest is EXAM, a

dummy variable that equals 1 if one's province has a curriculum-based provincial exam in the subject at the end of secondary school and zero otherwise. In the models predicting math and science scores, students in Alberta, British Columbia, Newfoundland and Quebec are assigned a 1 on EXAM. Students in Ontario, Saskatchewan, Manitoba, New Brunswick, Nova Scotia and the United States have a 0 on EXAM.

Table 6.1 presents our estimates of the impact of curriculum-based exams taken in 12th grade on test scores in 8th grade. Odd numbered columns present results for math. Even numbered columns present results for science. The first three rows present the results of simple regressions containing no controls for personal or school characteristics. American students score 8.6 percentage points below the Canadian sample in mathematics and 1.9 points below them in science. Row 2 tells us that students in provinces with exam systems score 8.3 points higher in math and 5.6 points higher in science. Adding controls for USA and for Francophone school (see row 3) slightly reduces the estimated impact of the exam to 7.5 points for math and 5.2 points for science. The coefficient on the U.S.A. dummy falls by about 37–40%, suggesting that exam systems are one of the reasons why Canadian students outperform American students. But the significant negative coefficients on the U.S.A. dummy in the mathematics equation suggest that there are also other reasons for the American lag in mathematics.

Now let us examine what happens in row 5 when controls are added for the demographic background of the student and the school's student body — books in the home, number of siblings, gender, languages usually spoken at home and school means for books at home and

Table 6.1.
Impact of Curriculum-Based Exams on Math and Science Achievement

	Subject exam		United States		French Speaking school		Adj R^2	
	Math	Science	Math	Science	Math	Science	Math	Science
Controls in model								
1. None			-0.086	-0.019			0.0076	0.0005
			(12.5)	(3.3)				
2. None	0.083	0.056					0.0256	0.0180
	(23.6)	(19.0)						
3. None	0.075	0.052	-0.052	-0.012	0.018	-0.049	0.0301	0.0284
	(19.9)	(17.1)	(7.4)	(1.9)	(4.6)	(14.6)		
4. Family demography – books in home, siblings, gender, home language	0.075	0.044	-0.033	0.002	0.048	-0.019	0.1143	0.1307
	(21.2)	(15.4)	(5.0)	(0.4)	(11.5)	(5.7)		
5. Add mean school demography books, languages	0.065	0.037	-0.010	0.015	0.108	0.025	0.1330	0.1431
	(18.2)	(12.8)	(1.5)	(2.6)	(21.3)	(6.2)		
6. Add TV, school mean TV, read for fun, home mathsci resources	0.055	0.030	-0.010	0.018	0.082	-0.002	0.1689	0.1887
	(15.7)	(10.5)	(1.5)	(3.2)	(15.5)	(0.5)		
7. Add student attitudes	0.056	0.032	-0.009	0.019	0.077	0.005	0.1887	0.1968
	(16.0)	(11.1)	(1.4)	(3.6)	(14.7)	(1.1)		
8. Add school resources and policies	0.045	0.031	-0.022	0.012	0.074	0.008	0.1977	0.2015
	(11.1)	(9.4)	(2.9)	(2.0)	(14.7)	(1.7)		
9. Add teaching strategies (12.6)	0.052	0.035	0.004	0.030	0.052	0.017	0.2230	0.2105
	(10.4)	(0.5)	(4.9)	(8.1)	(3.7)			

Notes: The provinces with curriculum-based external exams that counted as part of the course grade were Alberta, British Columbia, Newfoundland and Quebec.

proportion of students whose home language is different from the language of instruction. The estimated effects of exam systems on test scores remain strong: 6.5 percentage points for math and 3.7 points for science. The variables to be added to the regressions in models 6 through 9 — TV watching, home math-science resources and activities, attitudes, school resources and policies and teaching strategies — are hypothesized to be influenced by the existence of external exams. Consequently, row 5 presents our best estimate of the *total* impact (including indirect effects) of having a provincial exam in the subject at the end of secondary school on IAEP test scores at age 13. Provincial exams have very large effects: 24.3% of a U.S. standard deviation (about four-fifths of a U.S. grade level equivalent) in mathematics and 17.6% of a standard deviation (about two-thirds of a grade level equivalent) in science. **In a standard deviation metric, the impact of provincial exams on math scores is larger than the 22.6% of a standard deviation decline in Math SAT test scores between 1969 and 1980 and 2.9 times the magnitude of the increase in Math SAT scores since 1980.**

Adding the demographic variables also causes the U.S.A. dummy to fall to just about zero. This implies that the 8.6 point gap between the math scores of Americans and English speaking Canadians can be fully explained by a combination of demographics and exam system effects. On the other hand, the mathematics achievement of Americans and Canadians attending English speaking schools is way behind students from the same social class at French speaking schools in Canada. When one controls for the number of books in the home, the number of siblings, whether the language usually spoken at home is the same as the school's language of instruction, and school means for books and language, students at Francophone schools score 10.8 points higher in math and 2.5 points higher in science.

Row 6 of Table 6.1 presents the results of adding five to six measures of family behavior and math-science resources to the regression. Four of the additional variables — time spent watching TV, school mean on TV watching, a reading for fun index and school mean on a home resources for math and sciences index — appear in both the math and the science regressions. The additional variable in the science regression is a watching science programs on TV index. The mathematics regression has two additional variables: a dummy for owning a calculator and a dummy that equals one if the student says yes to: "Do you ever use a computer for school work or homework?" Adding these variables substantially increases R squares. The coefficients on EXAM fall by 15% in the mathematics regression and by 19% in the science regression. It would appear that induced changes in home behavior (particularly TV watching) account for an important part of the impact of provincial exams on test scores at age 13.

Adding student and parental attitudes to the model has no effect on the coefficient of EXAM, suggesting that the three attitude questions available in this data set do not mediate the effects of provincial exams (contrast rows 6 and 7). Similarly, adding teaching strategies to the model (contrast rows 8 and 9) fails to reduce the estimated effect of EXAM suggesting that teaching strategies also do not mediate the effects of provincial exams. Indeed, adding teaching strategies increases the estimated direct impact of EXAM. However, adding school resources and policies to the model (contrast row 7 and 8) reduces the EXAM coefficient in the mathematics regression by 20% of its original (0.056) level, suggesting that school resources and policies do mediate some of the effects of provincial exams.

When a full set of controls for teacher, administrator and parent behavior are included in the model, EXAM continues to have major effects. The provincial exams appear to raise percent correct by 5.2% (one-fifth of a U.S. standard deviation) in math and 3.6 points (17% of a U.S. standard deviation) in science. When we control for home characteristics, school resources and

policies and teaching strategies, American 13 year olds are equal in math to Canadians at English speaking schools but 4 points below Canadians at French speaking schools. In science Americans are equivalent to Canadians at French speaking schools and 3 points ahead of Canadians at English speaking schools. Estimating these models without the dummy variable for Francophone school does not change the estimated impacts of EXAM. In addition, models similar to row 5 were estimated using schools rather than students as observations. The coefficient on EXAM is 0.065 and the T statistic is 9.65.

Other Determinants of Mathematics and Science Achievement

Impacts of School Demography

Table 6.2 presents the full set of results when all school characteristics and teaching strategy variables are included in the regression. The average number of books in the homes of the students of the school has large effects. Controlling the actual number of books in the student's own home and doubling the school mean number of books at home increases math scores by 5.3 points and science scores by 3.5 points. Coming from a home where a different language is spoken than the language of instruction substantially lowers science scores but has little effect on math scores.

Students in the IAEP survey reported watching 14.7 hours of television a week. Heavy TV watching by the students at a school is apparently a sign of a school culture that does not encourage and support learning. If you attend a school where students average 20 rather than 15 hours of TV a week, your test scores are predicted to be 4.05 points lower and your science test scores are predicted to be 3.1 points lower. If you copy your friends and spend an extra 5 hours watching TV, your math score is predicted to go down a further 1.2 points and your science score is predicted to decline by 0.9 points.

Effectiveness of Nonpublic Schools

In their influential 1990 book, John Chubb and Terry Moe argued that the constraints placed on public schools by bureaucracy and democratic government make them inherently less effective than nonpublic schools that must compete for students and that are, thus, required to survive a market test. Clearly, however, their theory does not apply to Canada. When background characteristics of the students are controlled, students at nonpublic schools score 3.1–3.5 points lower on the IAEP science and math tests than public school students. Canadian nonpublic schools lag behind their public counterparts even though they receive considerable funding from provincial governments and local school districts.

As in the U.S., most nonpublic schools in Canada were started by religious denominations. Models were estimated which distinguished different types of nonpublic schools. When social class is controlled, public school students have significantly higher achievement in both math and science than students at religiously controlled schools. Students at secular nonpublic schools are slightly ahead in mathematics but significantly behind in science.

One is forced to conclude, therefore, that nonpublic schools are not inherently more effective at teaching math and science than public schools. My interpretation of these results is that

Table 6.2.
The Determinants of Math and Science Achievement at Age 13

	Standard deviation	Math percent correct coefficient	Math percent correct *t* statistic	Science percent correct coefficient	Science percent correct *t* statistic
School Demography					
Log average books in home	0.44	0.076	(13.39)	0.050	(10.33)
Speak different language in home	0.21	0.006	(0.59)	-0.045	(4.97)
Sci-math materials in home	0.65	-0.0081	(3.15)	-0.0042	(1.91)
TV hrs/wk avg in school	2.85	-0.0081	(11.62)	-0.0062	(10.85)
TV hrs/wk of student	8.65	-0.0024	(12.58)	-0.0017	(10.56)
Francophone school	0.44	0.052	(8.12)	0.018	(3.73)
Has curriculum-based exam	0.50/0.48	0.052	(12.63)	0.035	(10.40)
United States	0.25	0.004	(0.48)	0.032	(4.95)
Nonpublic school	0.40	-0.035	(7.89)	-0.031	(8.03)
School administrator behavior					
Specialist teachers in subject	0.50/0.50	0.016	(4.18)	0.006	(1.93)
Teachers studied subject in university	0.36/0.36	0.005	(1.05)	0.003	(0.83)
Specialized science laboratory	0.92	0.011	(5.70)	0.005	(3.54)
Tracking in 8th grade in subject	0.38/0.27	0.012	(2.60)	0.009	(1.79)
W/class ability group in subject	0.41/0.30	-0.017	(4.54)	0.001	(0.34)
Log size of 8th grade	0.93	-0.0054	(2.64)	-0.0052	(3.11)
Proportion of experienced teachers	0.22	0.019	(1.88)	0.008	(1.03)
Proportion of new teachers	0.14	-0.002	(0.10)	0.012	(0.84)
Class hours in subject	0.86/0.79	0.0025	(1.26)	0.0092	(5.10)
Teacher behavior					
Average homework hrs/wk in subject	0.62/0.44	0.0061	(2.37)	0.0044	(1.43)
Math difficulty index	0.41	-0.0042	(1.09)	-0.0024	(0.75)
Quiz index for subject	0.65/0.56	-0.0314	(11.46)	-0.0280	(10.94)
Listen to teacher index	0.77/0.70	-0.0360	(14.33)	-0.0030	(1.40)
Math group problem solving	0.96	-0.0226	(13.50)		
Math work alone on problems	0.70	0.0242	(10.62)		
Science do experiments ind.	0.86			0.0000	(0.05)
Science watch experiments	0.71			-0.0116	(5.77)
Science watch films index	0.66			-0.0111	(5.46)
Student behavior and attitudes					
Read for pleasure index	1.09	0.019	(12.49)	0.024	(19.25)
Have used computer for school work	0.49	0.025	(7.67)		
Have a calculator	0.31	0.018	(3.47)		
Watch science programs on TV	0.98			0.0156	(10.98)
Subject important to get job	0.63/0.86	0.0233	(8.66)	0.0181	(10.98)
Parents want me to do well in subject	0.61/0.95	0.0113	(4.16)	0.0106	(7.17)
Subject useful in everyday life	0.87/0.95	0.0314	(16.40)	-0.0103	(6.78)
Family demographics					
Female	0.50	-0.026	(8.16)	-0.058	(21.82)
Age	0.28	0.053	(9.80)	0.034	(7.45)
Number of siblings	1.36	-0.0107	(9.34)	-0.0110	(11.26)
Lang of instruct & home diff — 1	0.28	-0.014	(1.79)	-0.016	(2.49)
Lang of instruct & home diff — 2	0.22	0.003	(0.46)	-0.031	(5.21)
Less than 10 books at home	0.19	-0.098	(11.68)	-0.107	(15.10)
Between 11 and 25 books at home	0.32	-0.062	(11.81)	-0.060	(13.89)
Over 100 books at home	0.50	0.046	(13.10)	0.037	(12.75)
Adjusted R square		0.2230		0.2105	
Mean square error		0.2175		0.1780	
Number of observations		20,232		19,841	
Mean value of dependent variable		0.474		0.576	

religious schools probably place math and science education lower in their list of priorities than public schools. Their finances may also be more constrained. The religious denominations which control these schools and the parents who send their children to them are primarily seeking better discipline and ethical and moral climates that are consistent with their beliefs, not better math and science teaching. Fewer resources are devoted to math and science teaching and lower achievement results.

Effects of School Resources and Policies

The regressions yield some important insights into the school characteristics and policies that are associated with high achievement in mathematics and science. The school resource and school policy variables are based on a questionnaire filled out by the principal. A number of the variables constructed from this questionnaire appear to contain considerable measurement error. For example, student–teacher ratios calculated by dividing total enrollment by the size of the teaching staff range from 1.7 to 520 and have a standard deviation of 23. Estimates of the time during the school day that teachers have for preparing their lessons derived from class size and student–teacher ratios range from minus 15.8 to 0.93. Reported classroom hours of instruction in math and science also appear to contain measurement error. For the United States alone, the range is 2.08–6.67 hrs/wk in mathematics and 1.83–8.75 hrs/wk in science. One of the Canadian schools in the IAEP reportedly allocated only one hour a week to teaching science. The measurement error problems mean that coefficients on these variables are probably biased toward zero. It is also quite possible that school policies are in some cases jointly determined with average performance levels. Consequently one should view these coefficients as only suggestive of causal relationships, not structural estimates of a well-specified causal model.

Schools that have invested in specialized science laboratories and that use specialist teachers to teach the subject are significantly more effective at teaching science and mathematics. Tracking also appears to help, although the effects are small and only marginally significant. Within-class ability grouping of mathematics classes is associated with significantly lower math scores. Schools with smaller numbers of students in each grade appear to do better but the effect is small. A fourfold increase in the number of students in a grade lowers percent correct by less than one point. Having more experienced teachers has no significant effect on either mathematics or science scores. When teaching strategies are included in the model, the share of math and science teachers who had studied the subject at university has no significant effect on test scores. Such teachers appear to use more effective teaching strategies, however, because, when teaching strategies are not controlled, the proportion of math teachers who studied the subject at university has a statistically significant positive effect on math scores (not shown).

It is difficult to estimate the effect of time spent doing homework on learning in data sets where homework is not manipulated experimentally. Spending a lot of time doing homework might be due to a strong work ethic, strict parental supervision or a particularly demanding teacher. But it also might result from having difficulty doing the problems. Consequently, at the individual level, homework time will tend to be negatively correlated with unobservable ability, so coefficients will be biased. To minimize this problem, the specification uses school mean hours of homework rather than the time individuals spend on their homework to capture the effect of homework. Variations across schools in amount of time that is spent doing homework in particular subjects probably depends primarily on how demanding teachers choose to be and school policies regarding homework, not unobservable student ability.

The average number of classroom hours devoted to mathematics has no effect on math competence but assigning lots of homework does. For science, by contrast, devoting additional classroom hours to the subject has a big effect, while the effect of homework time is smaller and statistically insignificant. This pattern of results suggests that doing problems at home is an effective strategy for learning mathematics, but less effective for science. For science, an hour in the classroom has twice the effect on learning that an hour of homework has.

Teaching Strategies — Associations with Achievement

The conventional wisdom is that students spend too much time listening to teachers giving chalk and talk lessons or working alone at their desks. According to effective schools research, students do better on standardized tests when they attend schools that frequently test their own students. Science teachers are urged to use the discovery method — lots of experiments particularly experiments done by the student. Math teachers are urged to have their students solve mathematics problems in groups.

The IAEP students were asked how frequently — "every day" "several times a week" "once a week" "less than once a week" or "never" — they experienced different types of teaching. Both math and science students were asked about tests and quizzes and about listening to the teacher give lessons. The math students were also asked "How often do you solve problems in small groups during mathematics class?" and "How often do you do mathematics exercises by yourself during mathematics classes?" Students who took the science test were asked how often they watched science films, how often they watched the teacher do experiments and how often they did experiments themselves. The variables employed in the analysis were an average of the school mean response (with one-third weight) and the student's own response (with two-thirds weight). Other weighting schemes were tried including using the school means for these variables. This did not change the pattern of results discussed below.

These indicators of teaching strategy have strong relationships with test scores. Contrary to the effective schools research, students in classes where quizzes were common score below students in classes that tested less frequently. Students in classes where the teacher spent a lot of time giving lessons do more poorly in mathematics but not in science. Students who report doing lots of experiments do not perform better than students reporting few such opportunities. At least, however, they are probably not watching science films or watching the teacher perform experiments which is associated with students knowing less science. Spending a good deal of class time in group problem solving activities seems to lower the student's ability to solve problems on their own in a testing situation. Doing exercises alone, on the other hand, is associated with significantly higher performance on the IAEP test. This last finding is consistent with the earlier finding that math homework has a significant effect on IAEP test scores.

Interesting as these relationships may be, one must be cautious about interpreting them as reflecting causality. Longitudinal data on changes in test scores as students progress through school were not generated by the IAEP. The proxy for teaching strategy describes only the methods employed in 8th grade, not earlier grades. Consequently, these results do not tell us much about the effectiveness of different modes of teaching. Teachers may have been varying their strategies based on the background and ability of their students. The findings of effective schools research have received a lot of publicity. Many troubled schools may have adopted the frequent quizzes strategy and this may be the reason why quizzes have a negative relation with test scores. Similar arguments can probably be made about some of the other indicators of teaching strategy.

Effects of Student and Parental Behavior and Attitudes

All of the indicators of voluntary participation in school-like activities — reading for pleasure, watching science programs on TV, having a calculator and using a computer for school work or homework — have the expected positive effect on mathematics and science test scores.

Positive attitudes also had the anticipated effects. Students who report that their parents want them to do well in math or science indeed do better. Students who believe that math and science are important for getting a good job also do better on the IAEP tests. Students who strongly agree that "mathematics is useful for solving everyday problems" also do better. Surprisingly, those who make the same claim about science do not do better on the IAEP science test.

The Impact of Exam Systems on Parent, Student, Teacher and Administrator Behavior

The backwash effects of curriculum-based provincial exams on the behavior of 13 year olds, their parents, teachers and school administrators were examined by estimating models predicting these behaviors using schools as observations. The questionnaire completed by the principal provided the measures of school resources and policies. The student questionnaires provide our data on student, parent and teacher behavior. School means on each variable were calculated for the schools with at least nine students in the school sample and these were the dependent variables analyzed. Some of the schools selected to participate in the IAEP had considerably fewer than 30 age-eligible students. In developing the IAEP sampling frame, schools predicted to have only a few age-eligible students were combined into larger superschools for purposes of drawing the sample. When one of these schools was selected, the target sample of 30–34 students was distributed among the schools forming the superschool (IAEP, 1992c). Principal questionnaires were completed in each school, but sometimes the number of student interviews was too small to provide reliable estimates of school means. If the very small schools had been included, the estimated impacts of EXAM would have been slightly larger than the results shown in Table 6.3.

The specification was the same for all dependent variables. Nine variables were used: logarithm of the mean number of books in the home, the mean number of siblings, the proportion of the school's students whose home language was different from the language of instruction, a dummy for religiously controlled school, a dummy for secular nonpublic school, a dummy for French speaking school, a dummy for U.S.A. and EXAM. For outcome variables specifically associated with either math or science, EXAM's definition was the same as that used in the Table 6.2 regressions predicting test scores. When general school characteristics such as TV watching, total homework and class size were predicted, an average of the math and science EXAM variables was used.

The results are presented in Table 6.3. Each row represents a separate regression on data from 1366 to 1460 schools. The means and standard deviations across schools of each dependent variable are presented in columns 2 and 3. Coefficients on four of the variables — EXAM, U.S.A., French Speaking school and log books in home — appear in columns 4–8. Adjusted R squares appear in column 9.

Column 1 summarizes the hypotheses that were presented in the first section of this chapter. To the left of the slash, /, is the expected sign (based on *a priori* reasoning and the literature) of the impact of EXAM on this measure of home or school behavior. A question mark appears here if no hypothesis was generated for this variable. The +, – and 0s appearing to the right of the slash mark summarize the analysis of IAEP data presented in Table 6.2. A plus indicates that the variable had a significant positive effect on test scores at age 13. A minus implies a significant negative effect. A zero indicates no significant relationship.

Table 6.3.
Effect of Curriculum-Based Exams on Parent, Teacher and School Administrator Behavior

	Hyp.	Mean	StdDev School	Curric Exam coefficient	tstat	U.S. Coefficient	French Coefficient	Ln BookHm Coefficient	Adj R^2
Home behavior									
Tv Sch. avg. hrs/wk	-/-	14.7	2.85	-0.73	(5.0)	0.33	-2.1***	-3.4***	0.276
Read for fun index	-/+	1.85	0.28	0.05	(3.1)	-0.09**	0.10***	0.26***	0.144
Use computer for school work	?/+	0.40	0.24	-0.04	(2.6)	-0.06**	-0.12***	0.162***	0.160
Have calculator	?/+	0.88	0.13	0.04	(5.8)	0.04***	0.04***	0.067***	0.181
Watch science programs on TV	-/+	0.97	0.38	0.03	(1.5)	0.04	0.25***	-0.099***	0.110
P. want me to do well — math	+/+	2.53	0.22	0.05	(3.8)	0.05**	-0.02	0.030*	0.077
P. want me to do well — science	+/+	1.67	0.34	0.06	(3.0)	0.01	0.10***	0.172***	0.063
Parents talk about math class	+/?	0.62	0.17	0.05	(4.7)	0.08***	0.04**	0.023	0.039
P. talk about science class	+/?	0.47	0.17	0.06	(6.1)	0.06***	0.00	0.047***	0.048
Teacher behavior									
Total homework — hrs/wk	+/?	4.41	1.62	0.74	(8.3)	1.46***	-0.17	0.05	0.167
Math homework — hrs/wk	+/+	1.66	0.64	0.16	(4.2)	0.16**	0.008	0.124**	0.042
Science homework — hrs/wk	+/+	1.04	0.47	0.20	(7.0)	0.19***	-0.07*	0.072*	0.054
Emphasize whole number oper	-/-	1.68	0.49	-0.15	(5.1)	-0.12**	0.08*	-0.023	0.034
Math quiz index	+/-	1.62	0.52	0.27	(10.9)	0.53***	0.72***	-0.047	0.395
Science quiz index	+/-	0.89	0.38	0.12	(6.2)	0.68***	0.34***	-0.056**	0.331
Math group problem solving	-/-	1.48	0.62	-0.09	(2.6)	0.09	-0.60***	-0.107**	0.131
Math work alone on problems	+/+	3.22	0.37	0.01	(0.3)	-0.06	-0.13***	0.047	0.058
Science do experiments ind.	+/0	1.52	0.63	0.33	(9.6)	-0.11*	0.41***	-0.050	0.169
Science watch experiments	+/-	2.42	0.47	0.16	(6.0)	-0.08	0.24***	-0.121***	0.106
Science watch films index	?/-	0.94	0.48	0.02	(0.8)	0.40***	-0.06	0.006	0.059
Math listen to teacher	?/-	3.28	0.55	0.03	(1.3)	0.18***	-0.97***	-0.108***	0.587
Science listen to teacher	?/0	2.30	0.48	+0.04	(1.8)	0.38***	-0.60***	-0.057*	0.398
School administrator behavior									
Math specialist teachers	+/+	0.48	0.50	0.33	(13.1)	0.63***	0.13***	0.077***	0.237
Science specialist teachers	+/+	0.49	0.50	0.35	(13.7)	0.57***	0.00	0.122***	0.225
Took math courses in university	+/0	0.66	0.39	0.17	(7.9)	0.21***	-0.03	0.056*	0.113
Took science courses in university	+/0	0.69	0.38	0.20	(9.8)	0.15***	-0.17***	0.046*	0.173
Math class hours	+/0	3.97	0.89	0.19	(3.6)	0.04	0.34***	-0.303***	0.078
Science class hours	+/+	3.00	0.82	0.24	(5.4)	0.99***	-0.06	-0.017	0.174
Specialized science labs	+/+	1.95	0.95	0.63	(12.3)	0.33***	0.12	0.065	0.135
Tracking in 8th grade math	+/+	0.16	0.37	0.12	(6.3)	0.63***	0.08***	0.026	0.183
Tracking in 8th grade science	+/0	0.08	0.27	0.03	(2.2)	0.28***	0.11***	0.014	0.077
W/class ability group in math	?/-	0.23	0.42	-0.14	(5.7)	0.02	-0.09***	0.019	0.031
W/class ability group in science	?/0	0.10	0.30	-0.05	(2.9)	0.03	0.00	0.003	0.008
Propor. exper. teachers	+/0	0.59	0.24	0.02	(1.3)	-0.01	0.03	-0.035*	0.071
Propor. new teachers	-/0	0.16	0.15	-0.01	(1.4)	-0.03*	0.00	-0.002	0.049
Hours in school year	+/0	949	89	-7.4	(1.5)	36.7***	-11.0	5.80	0.029
Class size	-/+	24.8	6.1	2.1	(5.9)	2.1***	0.19	0.56	0.078
Teacher preparation time	+/+	0.31	0.27	0.07	(7.1)	0.07***	-0.02	-0.012	0.092
Log size 8th grade	?/-	4.15	1.03	0.75	(14.8)	1.29***	0.12	-0.025	0.280
Student attitudes									
Math important to get job	+/+	2.56	0.21	-0.01	(1.1)	-0.03	-0.05**	0.017	0.055
Science important to get job	+/+	1.93	0.33	-0.07	(3.8)	-0.13***	-0.21***	0.046*	0.131

Source: Regressions predicting the characteristics of 1366–1460 Canadian and American secondary schools. Control variables included but not shown were religious school, independent school, share of students whose home language was different from the language of instruction and mean number of siblings. Provinces with external exams included in final course grade were Alberta, British Columbia, Newfoundland and Quebec. Mean school char. based on grade 8.

Effects on Home Behavior

The hypotheses about the behavior of parents are strongly supported. As predicted in H-6 and H-7, students in provinces with exams watch 44 minutes less television a week, and are 5–6 percentage points more likely to report that their parents want them to do well in the examination subject and are also more likely to report that their parents have talked to them about what they are learning in school.

Opponents of externally set curriculum-based examinations predict that they will cause students to cut back on learning activities that do not have a direct relationship to the exams. This hypothesis, H-9, was operationalized by testing whether exam systems are associated with *less* reading for pleasure and *less* watching of science programs like NOVA and Nature. Neither of these hypotheses is supported. Indeed, point estimates of the effect of EXAM are positive (rather than the hypothesized negative) and the positive effect is statistically significant for reading for pleasure.

Effects on Teacher Behavior

It was hypothesized in H-11 that provincial exams would cause teachers to give more homework, to cover more cognitively difficult material, to schedule more quizzes and tests, to reduce the time that students spend doing group problem solving, increase the time that students work alone doing math problems and schedule more experiments in science class. H-14, which was derived from the writings of opponents of external examinations, makes opposite predictions: exams will induce teachers to focus on lower cognitive processes like computation in mathematics and facts and definitions in science. Students would have fewer opportunities to do experiments in science class.

All of the H-11 hypotheses tested in these data receive strong support. Provincial exams are associated with students doing 44 additional minutes of homework per week and 10–12 additional minutes per week of homework in mathematics and science. Students report taking more in-class quizzes and tests. Contrary to H-14 but consistent with H-11, emphasis on computation using whole numbers — a skill that should be learned by the end of 5th grade — declined significantly. In addition, students did more (not fewer) experiments in science class. Apparently, teachers subject to the pressure of a provincial exam four years in the future adopt strategies that are conventionally viewed as "best practice," not strategies designed to maximize scores on multiple-choice tests.

But they also apparently see through the conventional wisdom and allocate less time to group problem-solving activities. Also against conventional wisdom they may have also given students more in-class time to do problems on their own. This effect, however, is small and not significant.

Effects on School Resources and Policies

Most of the hypotheses, H-15, (a)–(k), about how administrator behavior would be affected by provincial exams are supported. The most striking effect is the big increase in the use of specialist teachers and teachers with university courses in the subject. The percentage of specialist teachers is 33 points higher in mathematics and 35 points higher in science. In provinces with exam systems, the proportion of teachers who took courses in their subject at university is 17 percentage points higher for math teachers and 20 percentage points higher for science teachers. The specialized science labs index is also two-thirds of a standard deviation higher. Classroom instruction hours are 5% higher in mathematics and 8% higher in science but no higher overall. This suggests that the increase in class time devoted to math and science comes at the expense of something else. In addition, tracking is more common in math and teachers get extra time to prepare their lessons.

The hypothesis that schools in exam provinces would try to hire more experienced teachers (in all subjects not just math and science) receives little support. They are slightly less likely to employ teachers with fewer than three years of tenure but they were also less likely to employ teachers with over 10 years of experience.

The other rejected hypothesis is the prediction that exam systems would stimulate reductions in class size. In fact, classrooms in provinces with exams have on average two additional pupils. Preliminary analyses of the determinants of test scores found a positive relationship between mean class size for the eighth grade and IAEP test scores. Could it be that principals "know" that large classes do not hurt student achievement on tests and that they obtain the resources necessary to hire more qualified teachers by increasing class size?

Hypothesis 16 that schools in exam system provinces would allocate resources in ways that our regression analysis suggests maximize IAEP test scores also receives strong support. Of the seven administrator behavior variables that had significant positive effects on IAEP test scores, all had significant positive relationships with EXAM.

I specified no hypothesis for how within-class ability grouping would be affected by external exams or how it would affect test scores. The cross-section regressions uncovered a negative relationship with IAEP test scores and with EXAM systems. Here again the signs of these relationships are, as predicted by hypothesis 16, the same.

I also specified no hypothesis about how school size would be affected by an exam system. The evidence indicates that schools in exam system provinces tend to be roughly twice the size of schools in provinces without exams.

Effects on Student Attitudes

Examination systems were not associated with a higher proportion of Canadian students thinking that mathematics or science were important for getting a job. These student attitudes apparently reflect the Canadian reality. Canadian employers apparently seldom use exam grades in hiring. Job applications were obtained from seven large companies located in Quebec, a province with a long tradition of exit exams. All of them requested information about degrees and certificates of skills but none requested information on grades in secondary school. School transcripts were asked for in a few cases, but interviewers reported that this was to confirm graduation not to screen on grades in school. These practices are not a consequence of legal prohibitions on requesting and using such information. A government approved official Canadian Manpower form obtained from the University of Montreal's College Placement office requests such information. Apparently, the availability of more reliable information on student performance in secondary school has not caused Canadian employers in Quebec (and presumably other provinces with examination systems) to ask applicants to provide information on secondary school grades.

Canadian provincial exams increase rewards for studying by signaling performance to students, parents, colleges and universities and making graduation dependent in part on externally assessed achievement; not by directly signaling achievement to employers.

How Does the U.S.A. Differ from Canada?

Column 6 of Table 6.3 presents estimates of the differences in student, teacher and administrator behavior between the U.S.A. and English speaking Canadian schools holding EXAM, school

control and social class constant. American 13 year olds do less reading for fun, are more likely to have a calculator, but less likely to use a computer for school work. American parents are more likely to talk with their children about what they are learning in math and science classes. American 13 year olds do 1.5 hours per week more homework than Canadian students. They take more quizzes, watch science films more and spend more time listening to teachers give lessons. Drill on simple arithmetic is less common.

American middle schools are much more likely to employ specialist mathematics and science teachers who have studied the subject in college. Science labs are better and an extra hour per week is typically devoted to teaching science. Tracking is much more common and the school year is 37 hours longer. Class sizes are larger and teacher preparation time is greater as well. Relative to Canada, anyway, the adults in the system appear to be trying hard to do the things that both conventional wisdom and Table 6.2 regressions tell us raise test scores. Why then do American students score significantly below Canadian students on the IAEP? The answer is demography (fewer books in the home and extra siblings) and the absence of curriculum-based external exams at the end of secondary school (see Table 6.1).

CHAPTER 7

CASE STUDY OF ENGLAND, SCOTLAND, FRANCE, THE NETHERLANDS AND THE UNITED STATES

Introduction

Despite similar standards of living, the secondary education systems of France, the Netherlands, Britain and the United States produce very different levels and patterns of achievement. In primary school, Americans do not lag behind their European counterparts. Reading ability varies little across these four countries. However, *when 14 year olds were compared at the beginning of the 1980s, the French and Dutch were about 1.3–1.5 grade level equivalents ahead of the Americans in math and science. At the end of secondary school, performance differentials appear to be even larger.*

What causes differences in secondary school achievement across these four nations? The next section describes these achievement differences. Seven hypothesized proximate causes of these differentials are evaluated in the following section. Four hypotheses can be rejected. The rest cannot: teacher quality, priority given to academics and student engagement and time on task.

The final section tackles the more fundamental question: "Why do American students, teachers, parents and school administrators place a lower priority on academic achievement than their counterparts abroad?" Why, for example, is student engagement in learning higher in France and the Netherlands? Some place the blame on culture, American anti-intellectualism, or historical tradition. Such *ad hoc* explanations cannot be ruled out (or in) by the analysis to follow. Our purpose is, instead, to propose an alternative explanation derived from the economic theory presented in Chapters 1 and 3.

Differentials in Academic Achievement

The differences in achievement levels at age 13, 14 and 15 during lower (first cycle) secondary school are summarized in Table 7.1. The table presents data from studies conducted in the 1980s and 1990s comparing France, the Netherlands, England, Scotland and the United States. The IEA studies sampled students at particular grade levels, not at particular ages. Consequently, age-adjusted scores on the IEA tests are reported where possible and information on the age of the sample is provided in the footnotes of the table.

If mean differences in achievement are to be given a grade level equivalent (GLE) interpretation, an assumption must be made about the relationship between grade level equivalents and the sample standard deviation for the test. This relationship varies across tests and across societies depending on the age of the students tested, the character of the test and the pace of instruc-

Table 7.1
Achievement in Lower Secondary School

	1991 IEA reading Age adjusted			1982 IEA Math Age 13–14 % correct	1983 IEA Science Age 14–15 (not adjusted for age) Mean (SD)		Adj. for Age Difference	1991 IAEP Mathematics Level Age 13 % correct Mean (SD)		Gain Age 9 to Age 13	1991 IAEP Science Level Age 13 % correct Mean (SD)		Gain Age 9 to Age 13
	Age 9	Age 14 Mean (SD)											
France	526	533	(68)	53.9				64.2	(20.3)		68.6	(17.1)	
Netherlands	494	523	(76)	57.1	63.7	(16.1)	62.2						
England				47.1	55.9	(15.7)	62.2	60.6	(21.4)	29.8	68.7	(17.5)	18.7
Scotland				48.4				60.6	(20.3)	26.5	67.9	(16.5)	20.8
United States	543	528	(85)	46.4	53.7	(16.7)	53.7	55.3	(20.9)	25.4	67.0	(16.4)	17.2

Columns 1, 2 and 3 are the age adjusted means and standard deviations of the overall reading score in the IEA reading study (Elley, 1992, pp. 108–9).
Column 4 is a weighted mean percent correct for students in the grade where the majority have attained 13:00–13:11 years by the middle of the school year from the Second International Mathematics Study (McKnight *et al.*, 1987, p. 124). The French, English, and American students all had the same mean age, 14.1. Mean age was 14.0 for Scotland and 14.4 for the Netherlands. Adjusting for the greater age of the Dutch students would have lowered their percent correct by about 2 points.
Columns 5 and 6 are the percent correct and standard deviation for 9th graders on the full 50-item IEA science test (Postlethwaite & Wiley, 1992, pp. 60, 74). An estimate of how U.S. students would have performed on the full test was made by subtracting 1.1 percentage points (the average difference between core and full test scores for England, the Netherlands) from the U.S.'s core test score. The mean age of students differed a great deal. mean age was 14.2 for England, 15.3 for the United States and 15.6 for the Netherlands.
Column 7 is an estimate of scores for the full 50-item IEA science test for students who are 15.3 years old, the mean age of U.S. students. The age gradient used was the average for Sweden (4.3) and Italy (7.4), the two countries for which it was available.
Columns 8, 9, 11 and 12 are the mean percent correct and standard deviation from the 1991 IAEP study of mathematics and science achievement of 13 year olds (IAEP, 1992a, b).
Columns 10 and 13 are the increase in the percent correct on items common to the tests given to 9 and 13 year olds.

tion. The approximate number of GLEs per SD for 13 year olds is about 4 for NAEP assessments, 3 for the Iowa Test of Basic Skills (8th graders) and 2.85 for the IEA science test. Where an estimate for the specific test is not available, I assume an SD on a test taken by 14 year olds equals 3 U.S. GLEs.

Reading

In the 1990/91 IEA study of reading, age adjusted scores indicate that American 9 year olds (see column 1 of Table 7.1) were reading about 58% of a U.S. standard deviation better than Dutch 9 year olds and about 0.20 SDs better than French 9 year olds. However, by age 14 differences between the countries (column 2) were tiny.

Mathematics

In the 1981/82 study of mathematics achievement of 13–14 year olds conducted by the International Association for the Evaluation of Educational Achievement (IEA), Dutch and French 13–14 year olds ranked number 2 and 3 (behind only Japan). Of the 17 industrialized nations participating in the study of 13–14 year olds, Americans were ranked 12th, English 11th and Scots 10th (McKnight *et al.*, 1987). After adjustment for small differences in mean age, American 14 year olds scored 10.7 (7.5) points below Dutch (French) students of comparable age (see column 4). The 1991 International Assessment of Educational Progress (IAEP) mathematics study obtained similar results (columns 8–10). The gap between French and American 13 year olds was 42.6% of a U.S. standard deviation (about 1.3 U.S. grade level equivalents). British students were about half-way between the French and the Americans (IAEP, 1992a). The gap remained roughly constant even though math achievement of 13 year old Americans improved by 0.20 SDs between 1982 and 1992 (NCES, 1994, p. 54).

The performance gap between the American and European students grows even larger during upper (2nd cycle) secondary school. Evidence of this can be found in Table 7.2. The Americans who participated in the Second International Math Study were high school seniors in college preparatory math courses like trigonometry, pre-calculus and calculus. This very select group, representing 13% of American 17–18 year olds, got 39.8% of the questions correct. The 6% of English students studying mathematics at A level got 59.8% correct (McKnight *et al.*, 1987). Substantial proportions of French and Dutch secondary students specialize in mathematics and science (20% of French youth are in the mathematics and science lines known as C, D or E of the *lycée général*) and the questions they are asked on their final examinations suggest that these students achieve at a very high level.

In 1964, performance gaps were also large, but at that time they were probably due almost entirely to the larger share of the age cohort tested in the United States.

Science

In the 1983 IEA study of science achievement of 14–15 year olds, the Netherlands ranked 3rd and the U.S. ranked last among 17 industrialized countries. After a rough adjustment for age differences, American students lagged slightly more than half a standard deviation (about 1.4 U.S. grade level equivalents) behind English and Dutch students (see column 5 of Table 7.1).

The 1991 IAEP science study found that at age 9 American students were ahead of students in Scotland, England and most other European countries (data for France and Netherlands are not available for this age). By age 13, English, Scottish and French students were ahead, although the differences were small and not statistically significant (IAEP, 1992b). The gap is smaller in the more recent study in part because overall science achievement of 13 year old Americans rose by 0.21 SDs between 1982 and 1992 (NCES, 1994, p. 56).

Few American upper secondary students study science in depth (see Table 7.2). Only 1 or 2% of the age cohort take two years of physics or two years of chemistry. Despite the highly select nature of this group (many of whom were taking the subject for Advanced Placement college credit), only 47.5% of the questions were answered correctly on the IEA physics exam and only 37.7% were correct on the IEA chemistry exam. The 4 or 5% of the age cohort of English youth who in their 13th year of schooling were studying these subjects for their A levels got 62.4% and 69.3% correct, respectively (Postlethwaite & Wiley, 1992).

In the 1971 IEA science study, 75% of the age cohort was assessed in the United States; only 13–29% of the age cohort was assessed in Britain, France and the Netherlands. This explains the performance gap between the U.S. and Europe at that time. While one cannot really be sure, the data suggest that the U.S. lag in mathematics and science at the end of secondary school developed during the 1970s. While standards were falling in the U.S., enrollment in rigorous academic programs was rising in Europe.

The Proximate Causes of Achievement Differentials

American elementary school students do not lag behind their counterparts in Europe. Indeed in reading they are substantially ahead and in science slightly ahead (see rows 1 and 13 of

Table 7.2
Achievement at the End of Upper Secondary School

	1982 IEA Mathematics			1983 IEA Science — Final Year of Upper Secondary School										1964 IEA Math		1971 IEA Science	
	Final Yr. Of Sec. Sch.			Physics			Chemistry			Biology			Total	College Math		All students	
	Percent correct	% Age group	% time Math	Percent correct	% Age group	Hrs/ week	Percent correct	% Age group	Hrs/ week	Percent correct	% Age group	Hrs/ week	science Hmwk	Percent correct	% Age group	Percent correct	% Age group
France														37.4	5	28.8	29
Netherlands														45.6	5	40.3	13
England	59.8	6	21	62.4	6	5.1	69.3	5	5.2	62.4	4	5.2	7.2	50.3	5	36.8	20
Scotland	42.8	18	17											36.4	5.4	36.8	17
United States	39.8	12	14	45.3	1	5.0	37.7	2	5.0	38.1	12	5.0	2.8	19.7 (18.0)	18	21.4 (15.1)	75

Column 1 is a weighted mean percent correct for students in the final year of secondary school from the Second International Mathematics Study (McKnight *et al.*, 1987, p. 124). The mean age was 17.8 for the U.S., 18.1 for England, 16.9 for Scotland. Column 2 is the share of the age cohort in advanced mathematics courses included in the study. Column 3 is the share of school time spent in mathematics classes.
Columns 4, 7 and 10 give the percent correct for students studying each science subject in the final year of secondary school.
Columns 5, 8 and 11 are the proportions of the age cohort taking each science subject in the final year of secondary school (for the U.S. it is the share of students taking their second year of the subject).
Columns 6, 9 and 12 are the number of hours per week spent in classes in each science subject. (Postlethwaite & Wiley, 1982, pp. 6, 36, 37, 69). The mean age was 17.5–17.10 for the U.S. and 18.0 for England.
Columns 13 and 14 are for the students specializing in mathematics in the final year of an academic secondary school. The British students are in their 13th year of schooling (Husen, 1967, vol 2, pp. 24, 118).
Columns 15 and 16 are for all academic (or comprehensive) secondary school seniors regardless of specialization (Comber & Keeves, 1973, p. 159).

Table 7.1). What then caused the large deficits in achievement in mathematics and science *at the end* of secondary school? Why does achievement lag in math and science and not in reading? Let us start by looking at seven proposed proximate causes of achievement differentials across countries.

(1) Diversity.
(2) Restricted access to secondary education.
(3) Teacher quality and salaries.
(4) Overall spending per pupil.
(5) Priority given to academic achievement.
(6) Time devoted to instruction and study.
(7) Engagement — effort per unit of scheduled time.

Our purpose is not to select a single most important explanation for the U.S. lag behind France and the Netherlands. Rather our objective is the more modest one of narrowing down the list of possible causes.

Diversity

Non-Hispanic whites score 0.45 grade level equivalents (GLE) higher than the overall U.S. average on NAEP reading tests, 0.56 GLE higher on NAEP mathematics tests and 0.98 GLE higher on NAEP science tests. If all French and Dutch students are compared to the 77% of American students who are neither black nor Hispanic, the European advantage is smaller. For mathematics at age 13, the gap would be about 0.9 grade level equivalents in both 1982 and 1991. In science, U.S. white 13 year olds were about one-half GLE behind the Dutch in 1983 and about 0.6 GLE ahead of French 13 year olds in 1991.

But, is it really fair to compare the non-Hispanic white population of the U.S. to the total population of France and the Netherlands? The U.S. is not the only country challenged by diversity. The Netherlands accepted 120,000 immigrants in 1990, twice the rate of immigration into the United States. The share of the students who are taught in a language different from their mother tongue is 6% in both France and the United States, 5% in Scotland, 12% in Canada, 15% in Northern Italy and 20% in Switzerland (IAEP, 1991a). If one is to adjust scores for the demographic and socioeconomic background of students, why not hold parents' education constant as well. If this were done, the French/Dutch lead over the U.S. would increase.

Access — Numbers of Students and Graduates

It is sometimes said that low achievement is the price one must pay for greater access. During the 1960s restricted access to academic upper secondary schools did appear to explain the lead of European upper secondary school students over their U.S. counterparts. By 1982, however, the situation had changed. Enrollment in rigorous academic programs had fallen in the U.S., while it was rising in France, Britain and the Netherlands. At present, only the United Kingdom exhibits the expected tradeoff between achievement levels and enrollment ratios (see Table 7.3). Only 43% of British 17 year olds and 12% of 18 year olds were attending secondary school full-time in 1991. Students preparing for A level exams achieve at high levels, but they represent a decided minority of the age cohort.

Table 7.3
1991 Enrollment and Completion Rates

	France	Nether-lands	United Kingdom	United States
Percent enrolled FT in secondary school[1]				
At age 16	92.0%	97.2%	62.4%	90.2%
At age 17	86.4%	90.0%	43.1%	74.7%
At age 18	57.2%	67.4%	12.3%	21.1%
At age 19	31.6%	41.5%	3.4%	5.0%
FT Equivalent enrollment in tertiary education[2]				
Age 18	19.1%	12.7%	24.4%	33.1%
Age 18–21	26.6%	19.5%	16.0%	33.4%
Age 22–25	12.7%	14.0%	4.8%	13.5%
Age 26–29	4.0%	4.0%	2.2%	6.2%
FTE years in school between age 16–29[3]	4.6 yrs	4.9 yrs	2.3 yrs	4.1 yrs
School enrollment rate age 5–29[4]	57.7%	55.2%	52.7%	55.2%
Secondary diplomas awarded/population of theoretical completion age[5]	75.8%	82.2%	74.4%	73.9%
First-degree graduates from universities/ Pop. of theoretical completion age[6]	16.3%	8.3%	18.4%	29.6%

[1]OECD, *Education at a Glance*, 1993, p. 117.
[2]OECD, 1993, pp. 129, 131.
[3]Calculated by summing the ratios of FTE enrollment to population for one year age groups from age 16 to 29 (OECD, 1993, pp. 117, 129, 131).
[4]OECD, 1993, p. 108.
[5]OECD, 1993, p. 176. The U.S. data do not include GED certificates. The labor market does not view the GED as equivalent to a high school diploma. GED certified high school equivalents are paid 6% more than high school dropouts but 8–11% less than high school graduates. The graduation rate for the UK is spuriously high because it counts regular GCSE exams taken at the end of 11th year of schooling as graduation. If one or more A level exams had been the definition of secondary school graduation, the graduation rate would have been 28% (Government Statistical Office, 1992, p. 8).
[6]OECD, 1993, p. 179.

French and Dutch youth, by contrast, now have higher enrollment rates than American youth. For example, 86.4% of French and 90% of Dutch 17 year olds were in secondary school in 1991 but only 74.7% of American 17 year olds. At age 18 enrollment in either secondary or tertiary education was 76% in France, 80% in the Netherlands and 54% in the United States. Despite lower college attendance rates in France and the Netherlands, larger shares of 18–21 year olds in France (52.2% on an FTE basis) and the Netherlands (56.4%) are enrolled in school (either secondary or tertiary) than in the United States (40.4%). Between age 16 and 29, the average American spends 4.1 FTE years in school, British youth spend 2.3 years, French youth spend 4.6 years and Dutch youth spend 4.9 years (OECD, 1993, pp. 117, 129, 131). These statistics contradict the widely held belief that the American education system, despite all its faults, at least achieves higher levels of participation than continental systems.

Not only are secondary school graduation standards higher, graduation rates are higher as well. In 1991 graduation rates were 82.2% in the Netherlands, 75.8% in France and 73.9% in the United States (OECD, 1993, p. 176). The large proportions of 18 and 19 year olds attending secondary school in France and the Netherlands indicate how high graduation standards are made compatible with high graduation rates. Students having difficulty with the fast-paced curriculum do not drop out; rather, they repeat grades and thus gain extra time to prepare for the demanding external exams. Many participate in vocational programs and apprenticeships that currently account for 54% of French and 70% of Dutch upper secondary students (OECD, 1993, p. 119).

The benefit of the early completion of secondary school in the U.S. is that large numbers of students enter tertiary education at a young age. However, some of the material covered during the first two years of college in the United States is covered in upper secondary school in France and the Netherlands. More bachelors degrees are awarded in the U.S., but there is doubt that the BAs awarded by America's second-rank universities represent the same standard of achievement as comparable European degrees. Hard evidence on this issue is not available.

Teacher Quality and Compensation

Quality

The quality of the people recruited into teaching is very important. The teacher's general academic ability and subject knowledge are the characteristics that most consistently predict student learning (Hanushek, 1971; Strauss & Sawyer, 1986; Ferguson, 1990; Ehrenberg & Brewer 1993; Monk, 1992).

Secondary school teaching does not pay well in the United States and it does not attract the kind of talent that is attracted into the profession in France and the Netherlands. Since university admission standards are higher in Europe, the university graduate pool from which European secondary school teachers are recruited is better educated on average than the college graduate pool out of which American teachers are recruited. Furthermore, American teachers are generally not the most talented members of the pool of college graduates. In 1977–78 the Math Scholastic Aptitude Test (SAT) score of intended education majors was 0.38 standard deviations (SDs) below the overall average, one SD below engineering majors and 1.2 SDs below majors in the physical sciences. The Verbal SAT of intended education majors was 0.30 SDs below the overall average (NCES, 1992, Table 124). In this respect, Britain is similar; entrants into programs preparing primary school teachers have significantly lower A level grades than average for university entrants (O'Leary, 1993).

In France, by contrast, secondary school teachers must do a double major in the two subjects for which they seek certification and then pass rigorous subject matter examinations. In 1991 only 31.3% of those who took the written exam for the *Certificat d'Aptitude au Professorat de l'Enseignement du Secondaire* (the most common of these examinations) passed it. The best teaching jobs go to those who pass an even more rigorous examination, the *Agrégation Externe*, which had a pass rate of 17.7% in 1991 (Ministère de l'Education Nationale et de la Culture, 1992, pp. 205, 206). French and Dutch secondary school teachers tend to be recruited from the middle (not the bottom half) of a pool of graduates of tertiary education which is in turn a selected sample of the nation's population.

Furthermore, American teachers are often not very expert in the fields they teach. Recent college graduates recruited into math or science teaching jobs spent only 30% of their college career taking science and mathematics courses. Since 46% had not taken a single calculus course, the prerequisite for most advanced mathematics courses, it appears that most of the math taken in college was reviewing high school mathematics (NCES, 1993b, pp. 428, 429). The graduates of the best American universities typically do not enter secondary school teaching because the pay and conditions of work are relatively poor.

Teacher compensation

The high academic standards for entry into upper secondary teaching in France and the Netherlands are sustainable only if wages and conditions of work are attractive. Data on the relative compensation of secondary school teachers are presented in rows 1 and 2 of Table 7.4.

American upper secondary teachers start at a wage that is 14% below that of the average worker and after 15 years of experience they earn only 33% more. Starting salaries are equally low in England. However, in France starting salaries are 6% above the all worker average and in the Netherlands they are 39% higher. *In France, England and Scotland upper secondary school teachers with 15 years of experience are paid 61–63% more than the average worker and in the Netherlands they are paid 132% more than the average worker.* For primary school teachers, by contrast, American pay levels are comparable to their Dutch and French counterparts (see Table 7.4, row 6).

Table 7.4
Teacher Compensation and Conditions of Work

	France	Netherlands	England	Scotland	United States
Compensation — teacher/all employees[1]					
Upper sec. teacher — start	1.06	1.39	0.87	0.91	0.86
Mid-career (15 yrs)	1.61	2.32	1.63	1.61	1.33
Lower sec. teacher — start	0.95	1.12	0.87	0.91	0.86
Mid-career (15 yrs)	1.44	1.58	1.63	1.61	1.33
Primary sch. teacher — start	0.93	0.97	0.87	0.91	0.84
Mid-career (15 yrs)	1.34	1.39	1.57	1.61	1.30
Teacher class contact hrs/yr[2]					
Upper secondary school	532	943	776	887	825
Lower secondary school	706	943	776	887	748
Primary school	875	1014	1013	950	1098
Class size[3]					
Upper secondary	29	24	16	15	25.6
Lower secondary	24	28	16	20	26.8
Primary	23	25	25	20	24.0
Sec. school students/teachers[4]	14.0	15.9	14.7	14.7	15.5
Sec. school expenditure/student relative to GDP per capita[5]	28.1%	24.7%	28.0%	28.0%	29.4%
Share of staff not classroom teachers[6]	36%	20%			47%

[1]Compensation of secondary teachers was calculated by multiplying their salary by the ratio of compensation to wages for manufacturing workers. This estimate of teacher compensation was then divided by average compensation of all workers. The figure for French upper secondary teachers is a weighted average of salaries for Agrege (20%) and others (80%) (Nelson & O'Brien, 1993, pp. 73, 74, 90, 91).

[2]Mean number of students in each class (Nelson & O'Brien, 1993, Table II.2).

[3]Mean number of hours of teaching a class per week times the mean number of weeks in the school year (Nelson & O'Brien, 1993, Tables II.3, II.4). Time devoted to preparation, in-service training and to nonteaching activities are not included in this total.

[4]The ratio of the number of full-time-equivalent pupils enrolled in public and private secondary schools to the number of full-time-equivalent secondary school teachers (OECD, 1993, p. 104).

[5]Data on expenditure relative to GDP per capita are from OECD, *Education at a Glance*, 1993, p. 95.

[6]Share of all staff employed in publicly funded elementary and secondary schools and ministries of education that are not classroom teachers. The nonteaching staff includes administrators at all levels, teachers aides, guidance counselors, librarians, nurses, custodial staff, food service workers, bus drivers, and clerical workers. The Dutch figure is for all three levels of schooling (OECD, 1993, p. 100). The French figure is for secondary education only (Ministère de l'Education Nationale et de la Culture, 1992, p. 184). The U.S. figure is for public elementary and secondary schools and does not include people working for State Departments of Education (NCES, 1992, p. 88). In the U.S. teachers' aides account for 8.8% of school staff.

The lower pay in the United States is not compensation for more attractive conditions of work (see rows 7–13 of Table 7.4). French upper secondary school teachers are in front of a classroom only 532 hours per year. Their American counterparts teach 825 hours per year. Teaching hours are similar to U.S. levels in England and Scotland (776 and 886 hours, respectively), but class sizes are substantially smaller. Dutch upper secondary teachers are the only group that clearly have heavier teaching loads than American teachers (Nelson & O'Brien, 1993).

When the salaries of college graduates are compared, those who enter teaching come out at the very bottom. Starting salaries of U.S. mathematics and physical science majors who entered teaching were 42% below the salaries of those who obtained computer programming and system analyst jobs and 35% below the starting salaries of those obtaining jobs in mathematics or physical science (NCES, 1993b, p. 26). University graduates who majored in physical science earned 78% more and economics majors earned 92% more than education majors over the course of their working lifetime (Kominski & Sutterlin, 1992). Since Americans with university training in mathematics and science can earn much more outside of teaching, those with talent in these areas are difficult to recruit into high school teaching. This results in most American teachers of mathematics and science being less well prepared than their counterparts in Northern Europe. This may help explain why American students lag behind French and Dutch students in mathematics and science, but not in reading. The fact that American primary school teachers are almost paid as much as French and Dutch teachers may also help explain why American 9–10 year olds compare favorably to their counterparts abroad.

Overall Spending per Pupil

Data on pupil–teacher ratios and spending per pupil are presented in rows 13 and 14 of Table 7.4. Pupil–teacher ratios are quite similar in the five countries, as is the ratio of spending per pupil to per capita GDP. Consequently, "low" overall levels of spending on K-12 education are not the cause of the lag in U.S. achievement.

Priority Given to Academics

If American spending per pupil is comparable to that in our four comparison countries, how come salaries are lower? What happens to the money saved by paying lower teacher salaries? Is it used to hire additional non teaching staff? Non teachers account for nearly one-half of the employees in K-12 education in the U.S., but only one-fifth of employees in the Netherlands and only 36% of secondary education employees in France (see the bottom row of Table 7.4). These staff perform services (such as bus transportation, sports activities, before and after school day care, counseling and occupational training) which are provided by other governmental organizations or the private sector in some other nations. The money also pays for the more attractive buildings, sports facilities, large school libraries, the numerous computers and colorful texts that are typical of American secondary schools. In part, this reflects the fact that books, computers and buildings are cheaper (relative to teachers of constant quality) in the United States. American spending patterns also reflect different goals. Academic achievement is the overarching goal (some would say the only goal) of French and Dutch secondary schools. In the U.S., academic achievement must compete with other goals. American schools are also expected to foster self-esteem, to provide counseling,

supervised extra curricular activities, musical training, health services, community entertainment (e.g., interscholastic sports), drivers' education and do all this in a racially integrated setting. These other goals require additional staff and different kinds of staff. They may not be served by hiring teachers with a strong background in calculus or chemistry, so resources get diverted from paying the high salaries necessary to recruit teachers thoroughly educated in chemistry. Unlike France, selection into teaching is not based almost solely on competence in the subject matter.

Time Devoted to Instruction

Many studies have found learning to be strongly related to time on task (Wiley, 1986). How do the five countries differ in the time that students spend in classrooms and doing homework? Table 7.5 reports the results of a variety of studies that compare time devoted to instruction. Estimates of total amount of time students in a country spend in school seems to depend on whom you ask and how the question is worded. The data quality problem was dealt with by

Table 7.5
Student Time — Instruction and Homework

	France	Netherlands	England	Scotland	United States
Total hours of instruction/year					
Primary sch — 1971[1]	918	1040	900	1040	900
5th grade in 1982[2]			984		1070
4th grade in 1991[5]	840	975			954
Secondary sch — 1971[1]	775	1120	900	1080	900
9th grade in 1982[2]		1007		1141	
8th grade in 1982[3]	1187	1000	896	1067	1008
9th grade in 1991[5]	1030	1092			792
Age 13 in 1991[4]	1073		960	1031	1003
Hours of homework in all subjects					
Hrs/wk — 8th grade in 1982[2]		8.4	6.0		9.6
Hrs/wk — 8th grade in 1982[3]	8	5	5	3	5
Hrs/wk — 12th grade math stud (1982)[3]					9
Hrs/wk — 12th grade sci stud (1982)[2]			11.5		9.8
Hrs/wk in 4th grade (1991)[5]	0.53 hrs	0.13 hrs			1.89 hrs
Hmwk GT 2 hrs/day — Age 13 (1991)[4]	55%		30%	15%	30%
Hrs/wk on language arts — 4th grade[5]	1.6	0.4			2.3
Time devoted to mathematics					
Math share — 8th grade (1982)[3]	12%	10%	13%	14%	14%
Hrs/wk math instr — Age 13 (1991)[4]	3.85		3.04	3.52	3.90
Hrs/wk math hmwk — Age 13 (1991)[4]	2.03		1.33	1.07	1.65
Hrs/wk math hmwk — 8th grade (1982)[3]	4.0	2.0	1.0	2.0	3.0
Time devoted to science					
Science share — 5th grade (1971)[1]	8%	2%	3%	3%	7%
Science share — 5th grade (1982)[2]			4%		10%
Science share — 9th grade (1971)[1]	8%	7%	8%	5%	10%
Science share — 9th grade (1982)[2]		25%	10%		20%
Hrs/wk science instr — Age 13 (1991)[4]	2.86		3.34	2.91	3.90
Hrs/wk science hmwk — Age 13 (1991)[4]	0.80		1.03	0.62	1.13

[1]First International Science Study (Passow, Noah, Eckstein & Mallea, 1976, pp. 262, 268).
[2]Second International Science Study (Postlethwaite & Wiley, 1992, pp. 14–33).
[3]Second International Mathematics Study (Robitaille & Garden, 1989, pp. 36, 79).
[4]International Assessment of Educational Progress (1992a, b).
[5]Lundberg & Linnakyla (1992, pp. 20–25, 57–59) analysis of the 1991 International Association for the Evaluation of Educational Achievement study of Reading.

averaging across studies. The total hours of instruction per year for each country was first expressed as a ratio to the U.S. level. Then a mean ratio was calculated by averaging the ratios from the studies that provided a comparison with the U.S. Sources are given in the footnotes of Table 7.5.

While estimates vary across studies, the pattern for secondary school students in the 1980s and 1990s is that French, Dutch and Scottish students spent 5–15% more time in school than U.S. students. English students, by contrast, spent 6–9% less time in school than U.S. secondary school students.

Differences in instruction time may explain some achievement differentials between countries. But they do not explain the generally poor showing of U.S. secondary school students in mathematics and science. While American students spend less total time in school, they get *more* mathematics and science instruction time than French, Dutch and Scottish students. Heavy European time commitments to foreign language study tend to crowd out mathematics and science instruction. In lower secondary school, all British students study one foreign language and French and Dutch students often study two. In America, by contrast, few lower secondary school students study a foreign language and, by the end of high school, graduates have taken an average of only 1.46 years of foreign language (NCES, 1992, p. 131).

European students learn mathematics and science more thoroughly than American students even when they spend less time on it. For example, in the IAEP study, mathematics instruction time was the same in France and the U.S., yet French students knew about 1.47 U.S. grade level equivalents more mathematics than American students. In science, by contrast, instruction time was one hour per week less in France, yet Americans still lagged about one-third of a U.S. grade level equivalent behind French students. Why does an hour of instruction in French and Dutch classrooms produce more learning than in American classrooms? Could heavier homework assignments be the explanation?

Homework

Cooper's (1989) meta-analysis of randomized experimental studies found that students assigned homework scored about one-half of a standard deviation higher on post tests than students not receiving homework assignments. The impact of homework on the rate at which middle school students learn was also significant, although somewhat smaller.

French lower secondary school students spent more time doing mathematics homework and homework of all types (see Table 7.5). In 1991, French 13 year olds averaged 2.03 hours of math homework a week compared to 1.65 hours in the United States, 1.33 hours in England and 1.07 hours in Scotland. This is consistent with their lead in mathematics achievement. In science, however, there is no evidence that Dutch and French students got more homework than American students. Furthermore, English and Scottish lower secondary school students do less homework (and have less instruction time) in mathematics and science than American students and yet outperform them.

Engagement — Effort per Unit of Schedule Time

Classroom observation studies reveal that American students actively engage in learning activities for only about half the time they are scheduled to be in a classroom. A study of

schools in Chicago found that public schools with high-achieving students averaged about 75% of class time for actual instruction; for schools with low-achieving students, the average was 51% of class time (Frederick, 1977). Overall, Frederick, Walberg and Rasher (1979) estimated that 46.5% of the potential learning time is low due to absence, lateness, and inattention.

Even more important is the intensity of the student's involvement in the process. The American high school teachers surveyed by Goodlad ranked "lack of student interest" as the most important problem in education. Sixty-two percent of 10th graders agree with the statement, "I don't like to do any more school work than I have to" (Longitudinal Survey of American Youth or LSAY, Q. AA37N).

Formal studies comparing ratios of on-task time to scheduled time are not available. Nevertheless, people who have visited classrooms in France or the Netherlands and the U.S. report that European teachers are less likely to be talking about extraneous matters and European students are more likely to be paying attention and doing what they have been assigned. My school visits in France and the Netherlands generated similar impressions.

Summary

Four of the seven proposed explanations for the U.S. lag in mathematics and science behind France, Britain and the Netherlands can be ruled out: diversity, restricted access, spending per pupil and time for instruction. The hypotheses that survive the first round of tests are lower quality teachers, lower priority attached to academic goals and lower levels of student engagement. With only 5 data points, no further narrowing of the list of hypothesized proximate causes is possible. Now let us look behind these proximate causes for ultimate causes. **WHY does an hour of instruction and homework time apparently have larger learning effects in England, France and the Netherlands than in America? WHY do French and Dutch secondary school mathematics and science teachers apparently expect more of their students than American teachers?** The next section of the paper proposes some tentative system level answers to our WHY questions. Our purpose is to show that a very simple application of the economic theory developed in Chapters 1 and 3 can provide a plausible explanation for the large system level differences in goals and learning efficiency demonstrated above.

Signaling As Ultimate Cause: External Examinations as Standard Setters

Examination Systems

When questions such as those placed in bold type above are put to French citizens and educators, they point to the high standards and pervasive influence of the *Baccalauréat*. In 1992, 71% of the age group took a *Baccalauréat (Bac)* exam. Fifty-one percent of the age group passed. Thirty-eight percent of the *Baccalauréats* awarded were *Bac Technologique* or *Bac Professionel* (i.e., in vocational lines) (Ministère de L'Education Nationale, 1993). This was a major accomplishment, for *Bac* exams set a very high standard. The three-year *lycée* programs that prepare 43% of the age cohort for the *Bac Général* are quite rigorous. *Bac* exams in mathematics, history/geography and French are set and marked by 23 regional *ac-*

adémies. School-based assessments are used for other subjects (Madeus Kellaghan, 1991, p. 17). The *Bac* exams taken in one's area of concentration are roughly comparable to the AP exams taken by American students seeking college credit for high school work. Cornell University, for example, generally awards advanced placement credit to recipients of the *Baccalaureate Général*.

The payoff to higher education is high, so access to university is highly prized. A *Bac* is necessary for university admission and the line pursued and the *mentions* obtained on the exam influence which university program one can enter. About 10% of those obtaining a *Bac Général* enter special programs which prepare them for the exam that regulates admission to the elite *Grandes Ecoles*. The job market also rewards young people who have passed the *Bac*. There are alternative lower level examined qualifications for employment such as the *Brevet d'Enseignment Professionnel* (BEP) and the *Certificat d'Aptitude Professionelle* (CAP), but the *Baccalauréat* confers greater access to preferred jobs. In 1987, unemployment rates for 15–24 year olds were 37% for those without a diploma, 22% for those with CAPs or BEPs, 18% for those with a *Bac* and 10% for university graduates (Ministère de l'Education, 1992b, p. 25).

Dutch university graduates earn 65% more than secondary school graduates at age 45–64 (OECD, 1992, 1993), so access to higher education is highly prized in the Netherlands as well. Examinations set by the Ministry of Education influence access to postsecondary education, so the high achievement of Dutch students in mathematics and science can be explained in the same way. The Ministry of Education sets an exam which has both essay and multiple choice components. The multiple choice component which represents half the written paper is graded centrally. The essay component is marked by the student's own teacher and by a teacher from another school with the aid of a marking scheme supplied by the Ministry. Oral components are administered by the student's teacher. In both France and the Netherlands, questions and answers are published in newspapers and available on video text. The published exams signal the standards that students and teachers must aim for.

Nine-tenths of English youth now take the General Certificate of Secondary Education (GCSE) exam at the end of 11th grade and an increasing number take A levels two years later. Scotland also has a system of external examinations. For the United Kingdom as a whole, the ratio of the number of school leavers passing at least one A level (or the Scottish equivalent) to the number of 19 year olds was 23% in 1991 (Government Statistical Service, 1993, p. 8). Completing an A level qualification lowers unemployment rates for 25–34 year olds from 16.9 to 6.9% and graduating from university lowers it further to 4.3%. Grades on the GCSE and A level exams are included on résumés and requested on job applications, so employment opportunities depend on school results as well (Raffe, 1984). University graduates earn 66% more than secondary school graduates at age 45–64 (OECD, 1992, p. 111). Performance on GCSE and A level examinations and the equivalent Scottish exams determines whether one can continue one's schooling and which university and program one can enter. There is a timing problem, however, because A level exam grades become available months after most students have applied to university. The solution that has been developed is quite ingenious. Universities base admissions offers on GCSE results and *teacher predictions of how the student will do in his or her A levels*. If the student does not achieve the A level grades that are predicted, the student may find that their admission to the specific university program is canceled. Even though teacher predictions of A level grades are decidedly optimistic, few school leavers suffer this fate. Nevertheless, the possibility is enough to keep British students studying hard up to the A level exam.

In the United States, by contrast, admission to the best colleges depends on teacher assess-

ments of *relative* performance — rank in class and grades — and *aptitude* tests that are **not** keyed to the courses taken in secondary school.

External assessments of achievement that directly affect access to preferred educational and job outcomes have clearly increased rewards for studying in these four European countries. They also change the *structure* of rewards for learning and, therefore, the incentive environment of students, teachers and administrators. These issues will be discussed under seven headings:

(1) Peer group norms.
(2) Teacher incentives.
(3) Administrator incentives.
(4) Competition among upper secondary schools.
(5) High standards in the external exams.
(6) *Redoublement*, grade repeating, as mastery learning and an incentive to study.
(7) Choice of specialization as goal setting.

Peer Group Norms

Peer pressure was discussed in my interviews of school staff and students in England, the Netherlands and France. The French educators I interviewed reported that peer pressure not to study occurred sometimes, but only in some lower secondary school classes, not at the *lycée* serving upper-middle class students that I visited. In lower secondary schools, the pressure appeared mild by American standards. In upper secondary schools particularly in the math-science line, the peer pressure was to excel. Discussions with Dutch and English students and educators produced similar observations.

Teacher Incentives

Most American secondary school teachers do not feel individually accountable for the **learning** of their students. The lack of accountability for learning stems from: (1) the rarity of examinations assessing student achievement in particular subjects relative to an external standard, and (2) the fact that most secondary school students receive instruction in a given subject from many teachers. Only coaches, band conductors and teachers of advanced placement classes are exceptions to this norm. They teach in environments where student achievement is visible to parents and colleagues and as a result feel accountable for outcomes.

In France and the Netherlands, by contrast, upper secondary students are in small classes that take most subjects together and generally remain intact for two or more years. Fewer than three teachers share responsibility for preparing each class for the external exams in the subject. In the Netherlands where schools are small, many subjects are taught by only one teacher. Since important rewards accrue to those who pass or do well in these exams, everyone takes them very seriously. The number of students taking and passing each exam is public knowledge within the school and among parents. Exam results influence teachers' reputations. Responding to such informal pressures, upper secondary school teachers strive to prepare their students for the external exam.

Administrator Incentives

In many European countries the record of each school in the external examinations (the numbers passing or getting high grades) is published in local and national newspapers. Recent reforms in England and Scotland, for example, have resulted in schools publishing annual reports that contain the grades received by last year's students in each examined subject. These reports are now sent to parents of current and prospective students. The school league tables have important effects on school reputations. Administrators seeking to strengthen their school's reputation are thus induced to give teaching effectiveness (as assessed by the external exam) first priority.

Competition among Upper Secondary Schools

For generations French and Dutch upper secondary schools have faced a competitive environment that is similar in many ways to the one faced by American colleges and universities. Funding has been on a per student basis, so schools experiencing an increase in applications have had an incentive to expand up to the capacity of their physical plant. Schools with strong reputations get more applications than they can accept and are, in effect, rewarded by being allowed to admit the "best" from their pool of applicants.

In the U.S. access to quality teaching and supportive peers depends on the parent's ability to buy or rent a home in a suburb with excellent schools. In France and the Netherlands access to the top upper secondary schools depends primarily on achievement in lower secondary school. This means that parents who want their child to attend the best upper secondary schools must make sure that their child studies hard in lower secondary school.

The Netherlands has three types of general secondary school — the VWO, the HAVO and the MAVO — and a system of lower vocational schools, LBO/LEAOs and KVBOs, which prepare students for both occupation specific and general education exams. The first year curriculum is supposed to be the same in all schools so that students can transfer between schools at its conclusion. In succeeding years, however, curricula and rigor diverge. Rigor and work loads are greatest at the six year VWOs, somewhat less demanding at the five year HAVOs and still less demanding in the four year MAVOs. These schools also differ in the foreign languages offered and the standard to which they are taught. The LBOs devote considerable time to occupationally specific curricula, so less time is available for general studies. Advice to parents about which type of school is appropriate for their child is based on the pupil's record in primary school and in some cases standardized tests as well (Nijhof & Streumer, 1988). Parents have the right, however, to select the type of school and which school of that type their child will enter. There are three parallel systems of education — a locally administered public system, a Catholic system and a Protestant system — so parents have a great deal of choice.

About a decade ago English and Scottish parents were given the right to send their children to schools outside the normal attendance area. Two years after choice became operational in Scotland, 9% of pupils entering secondary school nationally (11–14% in urban areas) attended a school outside their catchment area (Adler & Raab, 1988). Scottish parents who made this choice appeared to be behaving rationally for they tended to choose schools that had higher SES student bodies and were more effective than the school in their own catchment area. An analysis of school choice in the Fife Education Authority found that the schools chosen by those leaving their catchment area had better examination results than would have been predicted

given the pupil's primary school test scores and family background and the average SES of pupils at the school (Willms & Echols, 1993, Table 3). Consequently, the free choice of schools that prevails in our four European nations generates a competitive pressure on schools to excel that does not have any counterpart in the U.S. outside of cities with magnet schools.

Standards of the External Exam

External examinations at the end of secondary school are probably necessary if high achievement levels are to be attained, but they are not sufficient. Effects will be small if the exams are easy, are taken by only a small minority of students, or do not generate substantial rewards for successful students. British youth have lower achievement levels than French and Dutch youth. One possible explanation for this is that the passing standard of the GCSE is lower than for the *Bac* and the Dutch exams, and the more difficult A levels are taken by only a small minority.

High passing standards on external exams are clearly associated with high achievement levels. Does this reflect a cause and effect relationship? Yes, but causation runs both ways. High passing standards on medium- and high-stakes exams are politically sustainable only when most students taking the exam are able to meet or surpass the standard. At present, the median pupil in Britain in not expected to learn the entire multiplication table up to 10×10 until age 11. If the GCSE mathematics exams were made more demanding without strengthening mathematics teaching, failure rates might rise to politically unacceptable levels.

Does the passing standard also influence student effort? Yes it does. In High School and Beyond data, those taking more rigorous courses learned a good deal more between sophomore and senior year, even though their grade point average suffered as a result (Gamoran & Barends, 1987). Kulik and Kulik's meta-analysis (1984) of the educational literature found that students chosen to skip a grade or go to a compressed and accelerated curriculum score 75% of a standard deviation higher on tests (a few years later) than the matched nonaccelerated students. Repeating a grade effectively lowers learning goals and reduces the retained child's achievement a few years later by about 30% of a standard deviation (Holmes, 1989).

Over 100 experimental studies have been conducted on the effect of goal difficulty on various kinds of achievement. The effects are quite large: on highly complex tasks like school and college course work, specific hard goals raised achievement by 47% of a standard deviation (Wood, Mento & Locke, 1987). In the laboratory and field settings used by psychologists conducting this research, the subjects have generally accepted the goal set for them by the researcher. Achievement goes up, but the probability of failing to reach the goal rises as well. In most studies more than two-thirds of those in the "hard goal" condition failed to achieve their goal (Locke, 1968, pp. 163–165). Most of the studies examine behavior over relatively short periods of time. One would imagine, however, that if such experiments lasted a couple of years, those who consistently failed to achieve their goal might lower their goals or give up altogether.

Stedry (1960) found that when subjects who had already set their own goals were assigned even higher goals by the study director, they rejected the assigned goal and achievement did not rise. This appears to be what happens in American secondary schools. Most students reject the goals teachers set because the rewards for success are small. Others reject them because they appear unattainable.

How do European education systems induce students in upper secondary schools to set difficult learning goals and work toward them? They do not, as some have proposed for the

U.S., set a single high yea–nay standard that everyone is expected to meet. Young people are too different from each other for such a policy to work. External exams need to signal the level of a student's achievement, not just whether the exam was passed. Dutch external exams are graded on a 1–10 scale. Excellence in the *Baccalauréat* exams results in the award of a *Mention Très Bien*, an *Mention Bien* or a *Mention Assez Bien*. Once information on performance levels becomes available, employers and institutions of higher education will tend to base their selection decisions on it. Graduates with the strongest exam results have options not available to those with weak results, and the outcome is a system of graduated rewards. When the variance of achievement is high, incentives for effort are stronger on average under a graduated rewards system than under a single large reward attached to achieving a fixed standard (Kang, 1985).

The English GCSE and Scottish "Lowers" Examinations are taken by 90% of 16 year olds. They generate substantial and graduated rewards for learning what appears on the exams. Indeed the rewards for doing particularly well on these external exams appear larger than those in the Netherlands.

In the U.K., access to 6th form programs preparing for university, vocational technical programs of various kinds and employment depend on the student's performance in the GCSE and Scottish lowers. Since A level results are not available at the time initial university admission decisions are made, GCSE results influence which university and which field of study a student is admitted to. In the Netherlands the passing standard is high, but exceeding it by a large margin generates few rewards because the external exam results are only part of the student's overall grade and access to the most popular university fields of study is largely on a first-come first-serve basis. In addition, there is much less variation in the quality and reputation of Dutch universities than British universities.

Why then are English and Scottish 13 year olds assigned less homework than their American and Dutch counterparts? Why is their achievement in mathematics and science at age 13 significantly lower than in the Netherlands? As the time for the exam approaches in Britain, teacher demands and student effort increase substantially. At age 13, however, standards are low. **Why do the backwash effects of the secondary school graduation exams extend further back in the pupil's schooling in the Netherlands and France than in Britain?**

Redoublement as Mastery Learning and an Incentive to Study

One explanation for low British standards for 10–13 year olds is the lack of **immediate** rewards for doing well in classes. The external exams are three to six years away. Students are promoted to the next grade no matter how well they do in the previous grade. Those who fall behind inevitably slow the pace of the class in succeeding years. Primary school teachers do not feel accountable for how well students do in exams taken after four years of attendance at a secondary school. Secondary schools tend to be large and the teachers who handle the first year students also lack a sense of accountability for performance on exams that are more than three years in the future.

Social promotion of students is also the rule in the United States. A survey of teachers by Peter Hart Research Associates (1994) found that 46% reported pressure to "pass students on to the next grade who are not ready."

The situation is very different in France and the Netherlands. Pupils who fail more than one of their courses are generally required to *redoubler* or "to repeat the grade." In 1990 Dutch

redoublement rates were 7.5% per year in academic lower secondary schools, 5.1% per year in LBOs, the vocational lower secondary schools, and 13.3% per year in academic upper secondary schools (Central Bureau Voor De Statistiek, 1993, pp. 19, 20, 29). French rates of *redoublement* ranged from 6.8 to 11.0% per year during the four years of general lower secondary education, from 12.1 to 18.4% per year in the three year academic upper secondary schools and averaged 8.4% per year in the first two years of vocational upper secondary schools (Ministère de l'Education Nationale et de la Culture, 1992, p. 77, 93, 99). According to H. D. Lewis, the "basic motivation is to help the child himself, to ensure that the pupil is sufficiently well prepared so that he may fully benefit from work at a more demanding level (1985, p. 5)." For French teachers, *redoublement* is a form of Mastery Learning, a way of allowing some students extra time to achieve very demanding learning goals. Consequently, at age 19, 31.6% of French and 41.5% of Dutch youth are still in secondary school compared to 3.4% in Britain and 5% in the United States.

Redoublement is not inflicted only on children from lower class backgrounds. Often high aspirations can only be achieved by *redoublement*. In France selective upper secondary schools serving upper middle class communities have grade repeating rates that are nearly as high as schools serving lower income communities. For example, *Lycée Charlemagne*, an upper secondary school serving one of the richest neighborhoods in Paris, asked 14% of its entering class to repeat the year in 1992.

For French and Dutch teenagers, the threat of having to *redoubler* is a strong incentive to study. When I asked how the students who must *redouble* feel about it, I was told that they feel "**dishonored**". Since *redoublement* is a public event, parents also feel stigmatized, so they have an incentive to see that their child studies hard. In the Netherlands, students struggling with the fast-paced VWO or HAVO curricula are often given a choice: either repeat the year or transfer to a less demanding school. At the VWO I visited in the Netherlands, one-third of the entering class transfers to a HAVO or a less demanding VWO before the beginning of the third year. VWOs offer a fast-paced six-year university preparation program. Parents who want their child to enter a VWO are generally accommodated even when primary school teachers advise against it. The child's performance in school determines whether the parents' aspirations are realized or whether a transfer to a less demanding type of school is necessary. Being forced to transfer to a HAVO or a MAVO does not foreclose university attendance. With good grades at the end of the five-year HAVO program the student can transfer to a VWO, complete the final two years and then enter a university. In addition, numerous vocationally oriented higher education options are open to HAVO and MAVO graduates and transfers to university are feasible with good grades.

While other routes to university are possible, pupils who choose the fast track in 7th grade, a VWO, do not want to be forced "to get off the train". Students in the Netherlands and France are formed into classes which take most subjects together and which remain intact for two years and sometimes longer. Friendships tend to develop within this class. When I asked a Dutch student who, despite long hours of study, had been required to repeat a grade, why she had studied so hard, she responded **"I wanted to stay with my class!"** Students do not want to have to repeat the grade because it threatens to sever the friendships they have made in the class. Apparently, trying to keep up academically (i.e., accepting the academic goals of the school) is viewed positively by peers because it is an expression of commitment to the group. Those who refuse to study are apparently seen as rejecting the group. In these two countries peer pressure seems to encourage lagging students to study, not discourage them as in the U.S.

One would not expect the study effort of primary school pupils to be influenced by the

prospect of being retained. The hypothesis of significant threat-induced incentive effects applies to students in small secondary schools or large schools organized into small classes which take most subjects together and remain intact from year to year. Since most American students are in large high schools where peer relationships are not tied to taking particular courses, failing two courses does not sever peer relationships the way it does in Europe. Consequently, the threat of failing a course is not the powerful motivator that it appears to be in France and the Netherlands.

This would change only if something that really matters to youth — such as staying with age mates when they transfer to upper secondary school — were tied to passing a series of relatively rigorous courses or exams. This would produce a modified version of the "study hard or else you will be separated from your classmates" incentive that prevails in France and the Netherlands. The argument against retention is that it effectively lowers the learning goals being set for the student in subsequent years. Cross-section studies of primary school children indicate that retention reduces subsequent learning (Holmes, 1989). It also, apparently, increases future dropout risks (Grissom & Shepard, 1989).

Choice of Specialization as Goal Setting

All education systems give upper secondary students (and their parents) the right to select a specialty and the right to choose the rigor and difficulty level of either the school, the academic program or specific courses.

In France, four academic lines — literature and languages (A), economics and social science (B), mathematics and physical sciences (C), and biology (D) — have roughly equal numbers of students and together account for most of the *Baccalauréat Générales* awarded. the Mathematics-Physics-Chemistry line (C) is the most difficult, carries the greatest prestige and gives one the best chance of being admitted to a preparatory school for one of the elite *Grandes Ecoles*. Admission to the C line within a *lycée* is generally highly competitive. The Netherlands has a similar though less elaborate system of specialization within general upper secondary education. As in France the math-science line has the reputation of being the most difficult.

In France and the Netherlands, picking one's school and specialization effectively sets a specific learning goal. The prevalence of grade repeating and transfers to easier schools suggests that most students and parents initially set very difficult goals. The goal setting literature tells us that working toward a *specific* and *difficult* goal leads to greater effort and performance than being told to "do your best" or setting easy goals. Thus the French and Dutch pattern of setting highly ambitious goals maximizes average achievement levels even while it increases the number of students who fail to achieve the goal they initially set. **Why do French and Dutch parents select secondary schools and programs that are so challenging that many must repeat grades to keep up or transfer into easier programs and schools?** There are three reasons. First, the goal selected is visible to parents, relatives and neighbors and going for difficult goals confers prestige. Second, achieving difficult learning goals is rewarded by admission to preferred universities and fields of study and access to better jobs. Finally, the choice is generally made by the parent, not the child. Parents are better informed about the long-term benefits of achieving difficult goals and their own prestige rises when their child attends a selective school or pursues a difficult line of study. Parents may view the extra studying necessary in a rigorous specialty as a plus not a minus.

In America, by contrast, selecting difficult goals generates much weaker rewards. Everyone

in the neighborhood attends the same school. Students select individual courses, not programs or schools. Subjects are taught at vastly different levels, but the rigor of these courses is not well signaled to parents, relatives, neighbors, employers and colleges. In the past most colleges failed to factor the rigor of the courses taken into their admissions decisions. This is changing, however, and selective colleges now give preference to students who take Advanced Placement courses. Seniors applying to selective colleges are aware of this and often sign up for AP classes. However, grades on AP courses taken during the senior year become available long after admissions decisions are announced, so many seniors do not put the required energy into the course and skip the AP exam.

CHAPTER 8

SUMMARY AND LESSONS

This issue develops and tests a theory of student effort and school priority setting. The key assumptions of the theory are that:

- Learning requires the time and active engagement of the learner.
- Students have many other uses for their time, so learning is costly for them.
- The intensity of their investment in learning depends on a comparison of benefits — intrinsic and extrinsic rewards for learning — to costs. When the benefits of learning increase, student effort increases and learning increases.
- Parents, school administrators and teachers are also influenced by comparisons of the benefits of learning to the costs of focusing school resources and policies on academic achievement. When rewards for academic achievement grow for students and parents, pressures are generated that induce teachers, administrators and school boards to refocus their energies on raising academic achievement.

The first major prediction of the theory is that an increase in the extrinsic rewards for learning will cause student effort and achievement to increase. The primary extrinsic reward for achievement in high school is a higher probability of completing college. Thus the extrinsic rewards for learning in high school depend on the size of the payoff to college and on how contingent college admissions decisions are on achievement in high school. Time-series data suggests that changes in college selectivity and payoff may have contributed to the ups and downs of student achievement during the postwar period. The college payoff and test scores rose during the 1950s and peaked during the 1960s. Both series then declined during the 1970s, bottomed out around 1980, and then rose during the 1980s. The doubling of the payoff to college during the 1980s appears to have contributed to more homework being assigned and done and a big increase in the proportion of students taking rigorous mathematics and science courses. Analysis of cross-section data also supports a causal link. High school students living in communities where the payoff to college is large tend to take a heavier academic course load and are more likely to go to college.

The theory's second major prediction is that systems of regional or national curriculum-based external examinations that signal levels of achievement in specific subjects to colleges and employers will increase extrinsic rewards for learning and this, in turn, stimulates changes in priorities and behavior that result in higher achievement. The rest of the issue is devoted to testing this prediction.

The analysis indicates that externally set curriculum-based examinations at the end of secondary school have pervasive backwash effects on students, parents, teachers and school administrators. Nations with curriculum-based external examinations set higher minimum standards for entry into secondary teaching, pay their secondary school teachers more and focus school resources primarily on teaching. Mathematics and geography achievement is also higher in these nations. After Sweden eliminated upper secondary school exit exams during the 1970s, the proportion

of students taking rigorous college prep mathematics and science courses declined substantially.

New York, the only state in the U.S. with a system of curriculum-based external exams, ranks number 3 in teacher pay and the proportion of teachers with at least a master's degree and ranks number 2 in the share of the education dollar that is spent on instruction. The ratio of K-12 spending per pupil to public higher education spending per student is higher than in any other state. When adjustments are made for the ethnicity and parent's education of SAT test takers, New York State ranks higher on the SAT than any of the other 37 states where the SAT is taken by large numbers of students. New York State students scored 23% of a standard deviation higher than students of the same race and social background in other states.

Holding the social class background of students constant, students from Canadian provinces with examination systems were substantially (23% of a standard deviation) better prepared in mathematics and 18% of a standard deviation better prepared in science than students from provinces lacking such exams. The effect of an exam system on mathematics achievement of 13 year olds is larger in a standard deviation metric than the decline in math SAT scores between 1969 and 1980 that has been such a focus of public concern.

The analysis also found that examination systems had pervasive effects on school administrators, teachers and parents. In the provinces with external exams, schools were more likely to:

—employ specialist teachers of mathematics and science,
—have high-quality science laboratories,
—schedule extra hours of math and science instruction,
—assign more homework in math, in science and in other subjects,
—have students do or watch experiments in science class, and
—schedule frequent tests in math and science class.

At home students watched less TV and were more likely to report their parents wanted them to do well in math and science. In addition, parents were more likely to talk to their child about what they were learning at school.

There is no evidence that external exams caused any of the undesirable effects that opponents of external exams have predicted. Canadian students in provinces with exit exams did not watch fewer science programs on TV and reading for fun went up, not down as predicted. Contrary to the predictions of exit exam opponents, mathematics teachers decreased their emphasis on low-level skills like computation and science teachers arranged for students to do more, not fewer, experiments.

Only two of the testable hypotheses set forth in Chapter 3 were rejected. Jurisdictions with exit exams did not have smaller class sizes. Second, Canadian employers in Quebec were not using grades on high school exit exams in their hiring process.

Chapter 7 of the issue applies the theory to the task of interpreting differences between systems of secondary education in Britain, France, the Netherlands and the United States. In the Netherlands and France, learning in secondary school is assessed by difficult subject-specific external examinations and doing well on these exams generates very large rewards for the student. The reputations of teachers and schools are also affected by student achievement in these exams. Parents base their selection of the upper secondary school their child will attend and which academic or vocational program he/she will pursue, in part, on these reputations. Parents tend to set difficult goals for their children, so most students are placed in programs of study that for them are very demanding. Students are formed into classes which take all their subjects together, remain intact for two years or more and become the student's circle of friends. Students who are not progressing at the rate necessary to succeed on the external exam are asked either to switch to an easier curriculum or to repeat the year. Students do not want to be

forced to sever the friendships they have developed in the class, so they are strongly motivated to keep up with their studies.

In the United States, students are ranked relative to their classmates, not assessed against an external criterion, so they pressure each other not to study. Teachers are expected to pass almost all students, and if the class fails to study hard, the teacher is forced to lower the passing standard of the course. Subjects are taught at vastly different levels, but the rigor of the courses and the learning achievements that result are not well signaled to parents, neighbors, colleges and employers, so rewards for setting difficult goals are small.

The French and Dutch models of secondary education combine in one system many of the most drastic reforms that have been proposed for the United States:

(1) Externally set subject-specific achievement exams taken by almost all secondary school graduates which supplement not displace teacher assessment of students.

(2) Parental/student choice of upper secondary school with money following students.

(3) Mastery learning with teeth (those who fail two subjects in secondary school are required to either repeat the grade or transfer to a less demanding school or program).

(4) Secondary teaching is available only to those who demonstrate very high levels of competence in their subject. High entry standards are sustained by offering high wages and good working conditions.

(5) High standards for admission to the next stage of education.

This system of incentives and school organization appears to work for France and the Netherlands. A similar system, lacking only the externally set exit exams, also works well in undergraduate education in the United States. At the secondary level, however, such reforms are controversial. Successful implementation of any one of these reforms would be a major political undertaking. Implementation of the whole package of reforms is politically infeasible at present. Yet the analysis suggested that when, in Britain, just two elements of the package — mastery learning with teeth and attractive teacher salaries — are missing and a third element, school choice, is only recently introduced, achievement levels were substantially lower than in the Netherlands and France. Consequently, from a practical policy point of view, the message is not very positive. School climates and educational standards do not change rapidly and easily. France and the Netherlands have not discovered a cheap and painless route to high achievement.

The important lesson is that incentives — both their strength and structure — matter. There are less controversial ways of increasing the rewards for academic achievement. Reforms tailored to the American context have a greater chance of success than an effort to replicate the French or Dutch systems of secondary education.

Suggestions for Research

The effects of external examination systems must be assessed by comparing education systems, not by comparing individuals, classrooms or schools. Changes in examination systems take decades to have their full effect. This limits the number of independent observations and reduces the power of formal statistical tests. The low power of individual tests can be overcome by developing multiple tests of hypotheses.

This issue presented numerous tests of its central hypothesis, but it has not exhausted the possibilities. Other international data sets such as the IEA examinations need to be analyzed. When the Third International Mathematics Study is completed the number of countries for which changes in achievement over time can be analyzed will substantially increase. The theory

predicts that countries which have experienced large increases in the payoff to college or the selectivity of college admissions such as the United States and Britain will improve their position relative to other countries. It also predicts that countries that ended or weakened their curriculum-based external examinations about two decades ago — Sweden, Spain and Portugal, etc. — will find themselves falling behind. Comparative data on time on task, homework, student engagement and peer pressure against learning need to be collected and analyzed.

New York State should be compared to the other 49 states in additional data sets such as the NAEP state assessment, NLS-88 and High School and Beyond. Possible positive impacts of the ACT test and minimum competency tests on achievement also needs evaluation. The School Achievement Indicators Program in Canada is generating data on student achievement in Canada that will allow researchers to track over time the effects of reestablishing provincial examinations in Alberta and Manitoba.

The impacts of variations in the structure of external examination systems should be examined. The theory predicts that centrally administered Abitur examinations such as those given in Bavaria should create stronger incentives for study and lead to higher achievement than examinations where the student's teacher suggests the essay questions to be asked and grades the exam. Examination practices vary a great deal across Swiss cantons and this creates further opportunities for studying their impact. Exceptional nations such as Belgium that have high achievement levels despite an absence of curriculum-based external examinations need to be studied intensively.

More complex and sophisticated theoretical modeling will also be useful. The economic theories of greatest relevance to this complex system are agency theory, game theory, signaling theory and consumer demand theory, not production theory. This issue only scratches the surface. Deeper plowing of these furrows will, I predict, yield a large crop of new insights.

References

Adler, M., Raab, G. M. (1988). Exit choice and loyalty: the impact of parental choice on admissions to secondary schools in Edinburgh and Dundee. *Journal of Educational Policy*, **3**, 155–179.

American Federation of Teachers (1995). *Setting strong standards: AFT's criteria for judging the quality and usefulness of student achievement standards* (1990) (pp. 1–12). Washington, DC: American Federation of Teachers.

Association of Secondary Teachers of Ireland (1990). Information Sheet opposing changes in Examination Systems.

Bishop, J. (1977). The effect of public policies on the demand for higher education. *Journal of Human Resources*, **XII**, (3), Summer.

Bishop, J. (1985). *Preparing youth for employment*. Columbus, OH: The National Center for Research in Vocational Education, The Ohio State University.

Bishop, J. (1987). Information externalities and the social payoff to academic achievement. Center for Advanced Human Resource Studies Working Paper # 87–06. Ithaca, New York: Cornell University.

Bishop, J. H. (1989). Did the test score decline cause the productivity growth decline. *American Economic Review*, 178–197.

Bishop, J. (1989b). The productivity consequences of what is learned in high school. *Journal of Curriculum Studies*.

Bishop, J. (1990). Job performance, turnover and wage growth. *Journal of Labor Economics*, **8** (3), 363–386.

Bishop, J. (1990). The explosion of female college attendance. Center for Advanced Human Resources Discussion Paper 90–19.

Bishop, J. (1991). Achievement, test scores and relative wages. In M. Kosters (Ed.) *Workers and their wages* (pp. 146–181). Washington, DC: The AEI Press.

Bishop, J. (1993). The impact of academic competencies on wages, unemployment and job performance. In B. Malkiel & C. Plosser (Eds.) *Carnegie/Rochester Conference Series on Public Policy*, Vol.37, December.

Bradley, A. (1993). Not making the grade: Teacher firing spurs debate over standards and expectations for students. *Education Week*, **Sept. 13**, 1, 19–21.

Breland, H., Wilder, G. & Robertson, N.J. (1986). *Demographics, standards, and equity: Challenges in college admissions*. A report sponsored by American Association of Collegiate Registrars and Admissions Officers, The American College Testing Program, The College Board, Educational Testing Service and National Association of College Admissions Counselors, November.

Bureau of the Census (1975). *Historical statistics of the United States* (premium calculated from D913-D916, D920).
Bureau of National Affairs. (1986). Selection tests: Employer interest increasing. *Bulletin to Management*, **January 2**, 2, 7.
Central Bureau Voor De Statistiek. (1993). *Education statistics*. The Hague, Netherlands.
Chubb, J. & Moe, T. (1990). *Politics and markets and America's schools*. Washington, DC: The Brookings Institution.
Comber, L. C. & Keeves, J. P. (1973). *Science education in nineteen countries*. New York: John Wiley.
Commission on Workforce Quality and Labor Market Efficiency (1989). *Investing in people: A strategy to address America's workforce crisis*. Washington, DC: Department of Labor.
Competitiveness Policy Council (1993). *Reports of the subcouncils*, March. Washington, DC: Competitiveness Policy Council.
Cooper, H. M. (1989). *Homework*. White Plains, New York: Longman.
Costrell, R. (1994a). A simple model of educational standards. *The American Economic Review*, **84 (4)**, 956–971.
Costrell, R. (1994b). *Centralized vs. decentralized educational standards under pooling* (pp. 1–35) July 1994. Amherst MA: Department of Economics, University of Massachusetts.
Council of Ministers of Education, Canada (1993). *Report on the mathematics assessment* (pp. 1–106). Toronto, Ontario: Council of Ministers of Education, Canada.
Eckstein, M. & Noah, H. (1993). *Secondary school examinations*. New Haven: Yale University Press.
Ehrenberg, R. & Brewer, D. (1993). *Did teacher's race and verbal ability matter in the 1960s?:Coleman Revisited* (pp. 1–57). Ithaca, NY: Cornell University School of Industrial and Labor Relations.
Elley, W. (1992). *How in the world do students read?* The Hague, The Netherlands: International Association for the Evaluation of Educational Achievement.
Farr, R. & Leo F. (1982). Reading trend data in the United States: A mandate for caveats and caution. In G. Austin & H. Garber (Eds.). *The rise and fall of national test scores* (pp. 83–142). New York: Academic Press.
Ferguson, R. (1990). *Racial patterns in how school and teacher quality affect achievement and earnings*. Cambridge MA: Kennedy School of Government, Harvard University.
Finch, F. H. (1946). Enrollment increases and changes in the mental level of the high school population. *American Psychological Association*, June.
Fleming, J. & Chambers, B. (1983). *Teacher-made tests: Windows on the classroom*. San Francisco: Jossey Bass.
Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, **95 (1)**, 29–51.
Forsyth, R. A. (1987). (personal communication) Achievement trends for the Iowa tests of educational development in Iowa: 1942–1985. Iowa City: Iowa Testing Programs.
Frederick, W. C. (1977). The use of classroom time in high schools above or below the median reading score. *Urban Education*, **11 (4)**, 459–464.
Frederick, W., Walberg, H. & Rasher, S. (1979). Time, teacher comments, and achievement in urban high schools. *Journal of Educational Research*, **73 (2)**, 63–65.
Freeman, R. B. (1971). *The market for college-trained manpower: A study in the economics of career choice*. Cambridge, MA: Harvard University Press.
Freeman, R. (1975). Over investment in college training? *Journal of Human Resources*, **10**, 287–311.
Freeman, R. (1976a). A cobweb model of the supply and starting salary of new engineers. *Industrial and Labor Relations Review*, **29**, 236–248.
Freeman, R. (1976b). *The overeducated American*. New York: Academic Press.
Friedman, T. & Williams, B. (1982). Current use of tests for employment. In A. K. Wigdor and W. R. Gardner (Eds.) *Ability testing: uses, consequences, and controversies, Part II: Documentation Section* (pp. 999–169). Washington DC: National Academy Press.
Gamoran, A. & Barends, M. (1987). The effects of stratification in secondary schools: Synthesis of survey and ethnographic research. *Review of Education Research*, **57**, 415–435.
Gardner, J. A. (1987). *Transition from high school to postsecondary education: Analytical studies*. Columbus, OH: National Center for Research in Vocational, Ohio State University Research Foundation.
Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology*, **26**, 461–477.
Goodlad, J. (1983). *A place called school*. New York: McGraw-Hill.
Government Statistical Service (1993). *Education statistics for the United Kingdom: 1992*. London: Her Majesty's Statistical Office.
Graham, A. & Husted, T. (1993). Understanding state variation in SAT scores. *Economics of Education*, **12 (3)**, 197–202.
Griggs vs. *Duke Power Company* (1971). 3 FEP 175.
Grissom, J. B. & Shepard, L. A. (1989). Repeating and dropping out of school. In L. Shepard & M. L. Smith (Eds.) *Flunking grades: Research and policies on retention* (pp. 34–63). New York: Falmer Press.
Grubb, N. (1993). The varied economic returns to postsecondary education: New evidence from the national longitudinal study of the class of 1972. *Journal of Human Resources*, **28 (2)**, 365–382.
Gummere, R. (1943). The independent school and the post war world. *Independent School Bulletin*, 4, April (quoted in Arthur Powell, *Standards*, 1995, Chap. 6).
Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro-data. *American Economic Review*, **61 (2)**, 280–288.

P. D. Hart Research Associates (1995). *Valuable views: A public opinion research report on the views of AFT teachers on professional issues* (pp. 1–24). Washington, DC: American Federation of Teachers.

Hashimoto, M. & Yu, B. (1980). Specific capital, employment and wage rigidity. *Bell Journal of Economics*, **11 (2)**, 536–549.

Hieronymus, A. N., Lindquist, E. F. & Hoover, H. D. (1979). *Iowa test of basic skills: Manual for school administrators.* Chicago, IL: The Riverside Publishing Company.

Hollenbeck, K. & Smith B. (1984). *The influence of applicants' education and skills on employability assessments by employers.* Columbus, OH: National Center for Research in Vocational Education, Ohio State University.

Holmes, C. T. (1989). Grade level retention effects: A meta-analysis of research studies. In L. Shepard & M. L. Smith (Eds.) *Flunking grades* (pp. 16–33). New York: The Falmer Press.

Hunter, J. E., Crosson, J. J. & Friedman, D. H. (1985). *The validity of the Armed Services Vocational Aptitude Battery (ASVAB) for civilian and military job performance.* Washington, DC: Department of Defense.

Husen, T. (1967). *International study of achievement in mathematics.* New York: John Wiley & Sons.

International Assessment of Educational Progress (1992a). *Learning science.* Princeton, NJ: Educational Testing Service.

International Assessment of Educational Progress (1992b). *Learning mathematics.* Princeton, NJ: ETS.

International Assessment of Educational Progress (1992c). *Learning about the world.* Princeton, NJ: ETS.

International Assessment of Educational Progress (1992d). *IAEP technical report*, Volume 1. Princeton, NJ: ETS.

Jackson, G. (1988). Did college choice change during the seventies? *Economics of Education Review*, **7 (1)**, 15–27.

Jeff, (1991). Paper for ILR 360, October.

Jencks, C. & Crouse, J. (1982). Aptitude vs. achievement: Should we replace the SAT? *The Public Interest.*

Johnson, G. R. (1935). High school survey. *Public School Messenger*, **33**, 3–37.

Jones, S. (1993). *Reading but not reading well* (pp. 1–22). Ottawa Ontario, National Literacy Secretariat. (Stan Jones also kindly made unpublished data on literacy levels by age available to me.)

Kang, S. (1985). A formal model of school reward systems. In J. Bishop (Ed.) *Incentives, learning and employability.* Columbus, OH: National Center for Research in Vocational Education.

Katz, L. & Murphy, K. (1990). Changes in relative wages, 1963–1987: Supply and demand factors. *National Bureau of Economic Research*, **April**, 1–61.

Kominski, R. & Sutterlin, R. (1992). *What's its worth?, Educational background and economic status: Spring 1990.* U.S. Bureau of the Census, Current Population Reports, P-70, No. 32.

Koretz, D. (1986). *Trends in educational achievement.* Washington: Congressional Budget Office, April.

Kreeft, H. P. J. (1990). *Issues in public examinations.* International Association for Educational Assessment.

Kulik, J. & Kulik, C.-Li, (1984). Effects of accelerated instruction on students. *Review of Educational Research*, **54 (3)**, 409–425.

Lerner, B. (1990). Good news about American education. *Commentary*, **91 (3)**, March, 19–25.

Lewis, H. D. (1985). *The French education system.* New York: St Martin's Press.

Locke, E. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, **3**, 157–189.

Longitudinal Survey of American Youth (1992). *LSAY codebook.* Dekalb: Public Opinion Laboratory.

Lundberg, I. & Linnakyla, P. (1993). *Teaching reading around the world.* The Hague: International Association for the Evaluation of Educational Achievement.

Madeus, G. (1991). The effects of important tests on students: Implications for a national examination or system of examinations (pp. 1–19). American Educational Research Association Invitational Conference on Accountability as a State Reform Instrument, Washington, DC, June.

Madeus, G. & Kellaghan, T. (1991). Examination systems in the European Community: Implications for a national examination system in the United States (pp. 1–100). Contractor Report for the Office of Technology Assessment, U.S. Congress, Washington, DC, March 27.

Maier, M. H. & Grafton, F. (1981). Aptitude composites for the ASVAB 8, 9 and 10. Research Report 1308. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Marenbach, D. (1973). Rates of return to education in the United States from 1939 to 1959 (pp. 89–90). Stanford University Doctoral Dissertation, University Microfilms.

McDonald, J. (1994). The Canadian educational system. In R. Aronson & R. Thornton (Eds.) *The economic consequences of American education*, New York: JAI Press.

McKnight, C. C. *et al.* (1987). *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective.* Champaign, IL: Stipes Publishing Co.

Mento, A. J., Steel, R. P. & Karren, R. J. (1987). A meta-analytic study of the effects of goal setting on task performance: 1966–1984. *Organizational Behavior and Human Decision Processes*, **39**, 52–83.

Ministère de l'Education Nationale et de la Culture, (1992a). *Repères and References: Statistiques sur les enseignements et la formation*, Paris.

Ministère de l'Education Nationale et de la Culture (1992b). *L'Etat de L'Ecole*, Paris.

Monk, D. (1992). *Subject area preparation of secondary mathematics and science teachers and student achievement* (pp. 1–51). Ithaca, NY: Department of Education, Cornell University.

Mullis, I. V. S., Dossey, J. A., Campell, J. *et al.* (1994). *NAEP 1992 trends in academic progress*. Washington, DC: Department of Education, National Center for Education Statistics.

Murnane, R., Willett, J. & Levy, F. (1992). *The growing importance of cognitive skills in wage determination* (pp. 1–30). Graduate School of Education, Harvard University.

National Association of College Admission Counselors (1993). Reported in October 23, *Ithaca Journal*.

National Center for Educational Statistics (1992). *The Digest of education statistics: 1992*. Washington, DC: US Department of Education.

National Center for Educational Statistics (1993). *The condition of education: 1993*, Vol. 1. Washington DC: US Department of Education.

National Center for Educational Statistics (1993). *The condition of education 1994*, Vol. 1. Washington, DC: US Department of Education.

National Center for Educational Statistics (1993b). *Occupational and educational outcomes of recent college graduates 1 year after graduation: 1991*. NCES 93–162. Washington, DC: U.S. Department of Education.

National Education Goals Panel (1993). *The national education goals report: 1993*. U.S. Government Printing Office, Supt. of Documents.

National Federation of Independent Business (1987). Survey of NFIB membership using a questionnaire designed by John Bishop.

Nelson, H. (1990). *Memo to Albert Shankar on international comparison of expenditures on administration and teaching compensation*. American Federation of Teachers Research Department, March.

Nelson, H. & O'Brien, T. (1993). *How U.S. teachers measure up internationally: A comparative study of teacher pay, training and conditions of service*. Washington, DC: American Federation of Teachers.

Nijhof, W. J. & Streumer, J. N. (1988). The Netherlands. In *The encyclopedia of comparative education and national systems of education*. Oxford: Pergamon Press.

Noah, H. J. & Eckstein, M. A. (1988). Tradeoffs in examination policies: An international perspective. Paper presented at the Annual Meeting of the British Comparative and International Education Society, University of Bristol, September 15–17.

O'Leary, J. (1993). Universities reduce A-level score needed for degree courses. *The Times*, July 12.

Organization of Economic Co-operation and Development (1986). *Living conditions in OECD countries: A compendium of social indicators*. Social Policy Studies No. 3. Paris, France: Organization for Economic Co-operation and Development.

Organization of Economic Co-operation and Development (1991). *Switzerland: Reviews of national policies for education*. Paris, France: OECD.

Organization of Economic Co-operation and Development (1992). *Education at a glance*. Paris, France: OECD.

Organization of Economic Co-operation and Development (1993). *Education at a glance*. Paris, France: OECD.

Organization of Economic Co-operation and Development (1995). *Education at a glance*. Paris, France: OECD.

Passow, A. H., Noah, H., Eckstein, M. A. & Mallea, J. R. (1976). *The National Case Study: An empirical comparative study of twenty-one educational systems*. New York: John Wiley & Sons.

Postlethwaite, T. N. & Wiley, D. E. (1992). *Science achievement in twenty-three countries*. London: Pergamon Press.

Powell, A., Farrar, E. & Cohen, D. (1985). *The shopping mall high school*. New York: Houghton Mifflin.

Raffe, D. (1984). School attainment and the labor market. In D. Raffe (Ed.) *Fourteen to eighteen: The changing pattern of schooling in Scotland* (pp. 174–193). Aberdeen: Aberdeen University Press.

Rehn, G. (1980). Sweden. In *Education and youth employment in Sweden and Denmark*. Carnegie Council on Policy Studies in Higher Education.

Reynolds, A. J. (1992). Grade retention and school adjustment: An exploratory analysis. *Educational Evaluation and Policy Analysis*, **14 (92)**, 101–121.

Robitaille, D. & Garden, R. (1989). *The IEA study of mathematics II: Contexts and outcomes of school mathematics*. New York: Pergamon Press.

Roessel, F. P. (1937). Comparative mental ability of high school pupils in three Minnesota towns in 1920 and 1934. In *Minnesota Studies in Articulation* (pp. 122–128). University of Minnesota Press.

Rohwer, W. D. & Thomas, J. W. (1987). Domain specific knowledge, cognitive strategies, and impediments to educational reform. In M. Pressley (Ed.). *Cognitive strategy research*. New York: Springer-Verlag.

Rosenbaum, J. (1990). Do school achievements affect the early jobs of high school graduates? Results from the high school and beyond surveys in the US and Japan. Evanston, IL: Northwestern University.

Rundquist, E. A. (1936). Intelligence test scores and school marks of high school seniors in 1929 and 1934. *School and Society*, **43**, 301–04.

Shepard, L. A. & Smith, M. L. (1989). Introduction and overview. In L. Shepard & M. L. Smith (Eds.) *Flunking grades: Research and policies on retention* (pp. 16–33). New York: The Falmer Press.

Sizer, T. R. (1984). *Horace's compromise: The dilemma of the American high school*. Boston, MA: Houghton Mifflin.

Slater, C. & Hall, G. (1995). *County and city extra* (pp. 1–203). Lanham MD: Berman Press.

Stedry, A. C. (1960). *Budget control and cost behavior*, Englewood Cliffs, NJ: Prentice Hall.

Strauss, R. P. & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review*, **5 (1)**, 41–48.

Suskind, R. (1994). Put down, kicked around, honor students struggle on. *The Wall Street Journal*, May 26.

Swanson, E. O. (1973). Fashions in test scores, or what goes up must come down? *Student Counseling Bureau Review*, **25 (1)**.

Taubman, P. & Wales, T. (1972). Mental ability and higher educational attainment in the 20th century. National Bureau of Economic Research Occasional Paper 118.

Thomas, J. W. (1991). Expectations and effort: Course demands, students study practices and academic achievement. Paper presented at a 1991 Conference on Student Motivation sponsored by the Office of Educational Research and Improvement.

Tim (1986). Paper for ILR 360, October (pp. 1–6).

Tinkelman, S. N. (1966). Regents examinations in New York state after 100 years (pp. 1–15). Albany, NY: The University of the State of New York, The State Education Department.

Tuddenham, R. D. (1948). Soldier intelligence in world wars I and II. *American Psychologist*, **3**, 54–56.

U.S. Bureau of the Census (1993). Poverty status by state and ten large metropolitan areas in 1992. Table 25, Washington DC.

U.S. General Accounting Office (1993). *Educational testing: The Canadian experience with standards, examinations and assessments* (pp. 1–74). (written by Kathleen D. White). GAP/PEMD-93–11, Washington, DC.

The University of the State of New York (1965). *Regents examinations: 100 years of quality control in education* (pp. 1–27). Albany, NY: The State Department of Education.

Ward (1994). A day in the Life. *N.Y. Teacher*. January.

Wheeler, L. R. (1942). A comparative study of the intelligence of East Tennessee mountain children. *The Journal of Educational Psychology*, **33**, 321–334.

Wiley, D. E. (1976). Another hour, another day: Quantity of schooling, a potent path for policy. In W. H. Sewell, R. M. Hauser & D. L. Featherman (Eds.) *Schooling Achievement in American Society*. New York: Academic Press.

Willms, D. J. & Echols, F. (1993). Alert and inert clients: The Scottish experience of parental choice of schools. *Economics of Education Review*, **11 (4)**, 339–350.

Wood, R. E., Mento, A. & Locke, E. (1987). Task complexity as a moderator of goal effects: A metaanalysis. *Journal of Applied Psychology*, **72 (3)**, 416–425.

World Education News and Reviews (1993). (abstracting newsletter Faits Nouveaux), **6 (2)**, 6.