DOCUMENT RESUME

ED 340 775 TM 018 031

AUTHOR Madaus, George F.; Kellaghan, Thomas

TITLE Examination Systems in the European Community: Implications for a National Examination System in the United States.

SPONS AGENCY Congress of the U.S., Washington, D.C. Office of Technology Assessment.

PUB DATE Apr 91

NOTE 118p.; Contractor report prepared for the Office of Technology Assessment titled "Testing in American Schools: Asking the Right Questions." For related documents, see TM 018 025 and TM 018 037.

PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC05 Plus Postage.

DESCRIPTORS College Entrance Examinations; Educational Assessment; *Educational Change; Educational Policy; Elementary Secondary Education; Evaluation Methods; *Foreign Countries; *National Programs; *Public Schools; Standardized Tests; Student Evaluation; Testing Problems; *Testing Programs

IDENTIFIERS *European Community; United States

ABSTRACT

Whatever the cause of present-day problems in American education, there is little doubt that proposals for national curricula and examinations are a response to widely-held perceptions of serious problems in the schools. Although national examinations have been proposed, there has been little consideration of their purposes and structures. The evolution of testing policy in the United States is reviewed, and six proposals to establish national examinations in this country are considered. The second part of the paper describes the complexity, operation, and contexts of the examination systems of the member countries of the European Community, with particular attention to France, Germany, and the United Kingdom. The focus is on examination systems with the function of certifying students, rather than examinations that are national assessment systems similar to the National Assessment of Educational Progress. Whether or not national examinations would work in the United States as they do in Europe is not clear. European countries do not have external systems of examination outside the public school examination system, and teachers and government inspectors are an integral part of the school examination system. In the United States, a new infrastructure will be required for national examinations. Seven tables present comparative information about the European examination systems. An 185-item list of references is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED340775

EXAMINATION SYSTEMS IN THE EUROPEAN COMMUNITY: IMPLICATIONS FOR A NATIONAL EXAMINATION SYSTEM IN THE UNITED STATES[1]

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

George F. Madaus
Boisi Professor of Education and Public Policy
Director
Center for the Study of Testing, Evaluation, and Public Policy
Boston College

Thomas Kellaghan
Director
Educational Research Centre
St. Patricks College
Dublin Ireland

This contractor document was prepared for the OTA assessment entitled *Testing in American Schools: Asking the Right Questions*. This is being made available because it contains much useful information beyond that used in the OTA report. However, it is not endorsed by OTA, nor has it been reviewed by the Technology Assessment Board. References to this should cite the contractor, not OTA, as the author.

Paper prepared for the Science, Education, and Transportation Program, Office of Technology Assessment, U.S. Congress, Washington DC.

April, 1991

BEST COPY AVAILABLE

The idea that the United States needs to create a "national" examination system currently has broad appeal. Such a system is seen as essential to creating a world-class education system--one that would produce better-educated students, increase our nation's productivity, and restore our global competitiveness. Advocates advance several arguments. First, the United States is one of the few industrialized countries in the world without common national examinations. Second, the absence of such examinations is a key reason why we are outclassed in international comparisons of achievement. Third, a common national examination would help create and enforce a common national core curriculum. Fourth, national examinations would give teachers clear and meaningful standards to strive for and motivate students to work harder by rewarding success and having real consequences for failure. Fifth, the national examinations should incorporate "authentic" assessment techniques and measure higher-order thinking skills, not just the recall of facts. And sixth, instituting national examinations would help the United States overtake other nations in achievement and emerge the world leader in education.

In the advocacy of examinations, issues such as the validity of international comparisons of achievement, the extent to which external examinations affect motivation and achievement, and the link between school achievement and economic competitiveness are strangely regarded as unproblematic. Yet how, for example, does one explain the fact that for most of a century the United States was highly competitive in the absence of an external national examination system? While many people (in the business, academic, and government sectors) argue that the decline in America's economic

competitiveness is to be attributed mainly to the performance of the educational system (and, by implication, that reform of one part of the system will cure the country's economic ills), the evidence on which this argument is based is, at best, tenuous. It is difficult to see how correlations between the test scores of recent graduates and indices of productivity of the total workforce form a stronger basis for explaining decreased economic competitiveness than such factors as the amount of national investment in research and development, patterns of spending and saving in the community, concern with quality in production, human relations in industry, or an increase in self-gratification pursuits (including the use of drugs, alcohol, and tv) throughout the society (Kerr, 1991).

Whatever the cause of present-day problems in American education, there can be little doubt that proposals for national curricula and examinations represent a response to widely-held perceptions that schools are experiencing serious problems relating to discipline and the valuation of educational achievement. What is again strange, however, is that a rather simplistic approach within the educational system is being offered to solve problems which are exceedingly complex and have deep roots in our society. Without necessarily sharing the view that patterns of academic performance are immutably moulded by social and economic conditions outside the school, one can appreciate that schools and students have not escaped the influences, sometimes positive but often negative, of major demographic, technological, economic, and social-psychological changes in our society in recent years (Coleman, 1990; Coleman & Husen, 1985). Given the pervasiveness of such influences, one can

hardly expect that matters will be put right simply by making adjustments, however radical, within the educational system, though such a view would not be without precedent in the history of educational reform (see Madaus, Airasian & Kellaghan, 1980). Recently, Harold Howe 2nd (1990) has pointed to what he calls the "basically erroneous assumption of most school-reform initiatives: the belief that schools can repair all the damage done by inadequate family attention and limited community services," pointing out that families and communities, as well as a variety of other sources of learning, are as important, maybe more important, than schools in educating the young.

Even if we confine our attention in reform to the school, there are many aspects of the educational system which invite attention, such as instructional techniques, length of school day and of school year (see Table 1 for data on lengths of school year in European countries), the quality of preservice and inservice courses for teachers, and a variety of methods of quality control.

As for the particular reforms which are being considered--common national curriculum[2] and national tests--there has been very little consideration of what the purpose of the examinations would be, what they would look like, at what age they would be administered, who would control them, or how much they would cost. Nor have the positive and negative consequences, short-term and long-term, on examinations driving teaching and learning been evaluated. While positive effects of examinations or teacher and student motivation are invariably stressed, the negative effects, particularly on low-achieving students, do not seem to have been considered. Further, there has

been no discussion about how a new testing program would fit into current extensive national, state, and local testing programs.

We shall not deal further with these broad issues though we regard them as forming an important context for a consideration of the role of examinations in educational reform. Our focus shall be on examinations, outlining first of all current proposals relating to national examinations in the United States, and then describing examination systems in Europe which some commentators suggest should act as a model for an American system.

Indeed, rhetoric about why the United States needs national examinations often includes claims about European examination systems. However, many of the proponents of examinations do not seem to have an accurate understanding of European systems. In particular, there often seem to be misunderstandings of the complexity and uniqueness of examination systems in Europe, as well as a lack of knowledge of why and how such systems developed as they did. Moreover, there is often a failure to consider the broader educational and social context in which European examinations are embedded, how they establish different patterns of power and authority in different countries, fit into the overall educational infrastructure of a country, and how the very examination technology and its infrastructure have evolved differently in different countries over many decades.

In the first part of the paper, we describe the evolution of attitudes that has brought us to the stage of acceptability of national examinations, national standards, and a national curriculum. We then review the evolution of testing policy in the United States over the last

30 years and how that development has shaped the acceptance of the concept of a national examination. After that, we describe six proposals to establish national examinations in this country.

In the second part of the paper, we attempt to describe the complexity, operation, and contexts of the various examination systems of the member countries of the European Community (EC), paying particular attention to France, Germany, and the United Kingdom. More specifically, we describe important features of examination systems as they developed in Europe. Then we describe the educational systems in which examinations operate since the examinations are closely tied to the structure of systems. Finally, we describe the examinations themselves, considering the tradition of examinations, changes in examination systems, and the models that operate in different countries. We concentrate on those examination systems whose function it is to certify students rather than on national-assessment systems similar to NAEP.

THE EVOLUTION OF ATTITUDES TOWARDS A NATIONAL CURRICULUM AND NATIONAL TESTS IN THE UNITED STATES

Changing Attitudes Toward Local Control

Any discussion or eventual adoption of a national test must contend with strongly held traditional beliefs in the United States in the sanctity of local control of education and with the de facto loci of control that have evolved since the 1950s. In the past, opposition to the concept of a national test, national educational standards, and a

national curriculum were partly rooted in the belief in the sanctity of local control of education. For example, in 1962, the Executive Committee of the American Association of School Administrators (AASA) felt it necessary to reaffirm "its belief in the principle long held in the United States that education is a state function, and the principle works exceedingly well when the responsibility of the state is delegated in large measure to local boards of education" (quoted in Campbell & Scroufe, 1965, pp.2-3). Similarly, the early 1960s saw a proposal for establishing a national commission for curriculum research and evaluation because of the "growing awareness of the need to improve American education, particularly the core curriculum" (Hanna, 1961, p. 337). However, to guard against infringing on local and state control, the proposed commission would have been advisory and would have had no administrative mechanism to enforce its proposals.

A wariness of federal control at the expense of local and state control continued during the 70s. For example, in 1977, Secretary of Health, Education and Welfare Joseph Califano, responding to President Carter's partiality towards Senator Pell's idea for an optional national test, responded that "Any set of test questions that the federal government prescribed should surely be suspect as a first step toward a national curriculum...In its most extreme form, national control of curriculum is a form of national control of ideas" (quoted in Smith, O'Day & Cohen, 1990, p.10).

As late as the beginning of the 1980s, state-mandated tests were "often viewed as an unnecessary intrusion by the state into local affairs" (Anderson & Pipho, 1984, p.209). However, by the mid-

eighties, state testing had become the preferred means of trying to effect change in education (Anderson & Pipho, 1984; Madaus, 1985). By 1988, the climate of the 70s described by Califano had changed sufficiently to permit the 100th Congress to enact Senator Pell's idea of an optional national test (PL 100-297). Currently, there are at least half a dozen proposals for some sort of national test.

While the belief in local control has been greatly eroded, it is by no means dead (Gallup, 1986). And, while state control has become a reality, federal control is still anathema to many Americans. The National Center on Education and the Economy (1989), in describing the federal role in education, felt the need to reassure America that:

> the federal government does not propose to take over responsibility for education. America does not want and will not tolerate uniform federal standards for education. The states have the primary role in setting education policy (p. 10).

Therefore, the distinction between federal and national tests and curricula becomes increasingly important to their viability. When discussion about a national exam and a national curriculum is replaced by realization that such a system will actually be implemented, and certainly affect local and state control, the debate over the merit of local, state, national, or federal control is sure to re-emerge (see Finn, 1991 for one view on the future of local control and Cuban, 1990 for a discussion of four views on local control and national standards). For example, in January 1990, the American Association of School Administrators (AASA), which represents 18,000 local school superintendents, considered the issue of a national test and "The idea of a single national test of 40 million schoolchildren was rejected

though not federal, curriculum (stripped of the monstrous camel metaphor and the derogatory word dictated) is increasingly accepted today (Commission on the Skills of the American Workforce, 1990; Shanker, 1990, 1991; Shanker, 1990; Smith, O'Day & Cohen, 1990).

To assuage critics, NAEP's proponents characterized the operation as being different from mounting a nationwide testing program. Further, a non-governmental body was put in control[3] and procedural and technical safeguards were established that prohibited any individual, school, district, or state from being singled out for description or comparison (Tyler, 1965). NAEP was born and can be regarded as the nation's first national testing program. We shall describe below current changes being made in the original design of NAEP that alter the original safeguards that made its birth possible.

A test can be truly national in that it is administered country wide but not necessarily federal in its locus of control (Smith, O'Day & Cohen, 1990). For example, the Scholastic Aptitude Test (SAT) of the College Entrance Examination Board (CEEB), albeit private, is a national but not federal testing program and, as we shall see, several current proposals envision a national, rather than a federal, testing program. NAEP is an interesting hybrid between a national and a federal test; initially sponsored by the Carnegie Corporation, it has been under the control of a non-governmental body since its inception but is federally funded through the Department of Education. One could argue that purse strings are the ultimate form of control. Therefore, in considering proposals for a national test, governance issues (who is eligible to serve on a board, who appoints members, for how long, what is their function?, etc.), the power of the sponsoring

outright" (quoting Bruce Hunter, AASA's executive director, in De Witt, 1991 p. E-5). The influence of a powerful national exam system on local and state control needs further discussion and will probably continue to be an issue for the foreseeable future.

Changing Attitudes Toward the National Assessment of Educational Progress

The early 1960s also saw the birth of the idea of developing the National Assessment of Educational Progress (NAEP). When NAEP was first proposed, it was envisioned that it would provide a needed indicator of educational progress for the country analogous to the Gross National Product (Tyler, 1965). Critics immediately charged that whoever decided what goes into such tests would determine what would be taught in the schools, and would eventually lead to a federally-controlled curriculum. For example, Harold Hand in an article entitled National Assessment viewed as the camel's nose offered the following arguments against a national test: it would set obstacles to the realization of the goal of equality of educational opportunity; it would be the nose under the tent which would be followed by a monstrous camel in the form of a centrally controlled, dictated curriculum; it would stultify the curriculum; it would stultify local curriculum innovation and experimentation; and it would encourage cheating on the part of students and teachers alike (Hand, 1965, p.9; see also Beymer, 1966; Hand, 1966; Saylor, 1970).

To date, none of Hand's fears has proved true regarding NAEP. However, all of them could equally well be applied to the national tests that are being discussed today. In fact, his second vision of a central,

body to compel students to take the test, whether it is optional or mandatory, the purpose of the tests, use of results and the source of financing (federal, state, private foundation, or fee paying) become important.

Public Attitudes Toward National Tests and a National Curriculum

The Annual Gallup Poll of the Public's Attitude Toward the Public Schools offers an indication of how attitudes toward a national test and a national curriculum have evolved over the last two decades. Since 1970 Gallup has asked respondents:

> Would you like to see the students in the local school be given national tests, so that their educational achievement could be compared with students in other communities? (Gallup & Elam, 1988, p.41).

Since 1970, a majority of people have favored such a test with sentiment strongest in 1988. (The question was not asked in the 1990 survey.) However, the meaning of the term "national tests" in the Gallup question is far from clear. It might well be interpreted to mean a traditional commercially-available nationally norm-referenced achievement battery given by most local school districts with which most parents are familiar.

However the following question removes any such ambiguity and speaks directly to several current proposals for a national test:

> Should all high school students in the United States be required to pass a standard nationwide examination in order to get a high school diploma? (Gallup & Elam, 1988, p. 40).

While one out of two respondents supported a national high-school certification exam in 1958, the ratio had increased to almost 3 out of 4 the last time the question was asked in 1988. The question is interesting because of the phrases "all high-school students" and "a standard nation-wide examination". These imply a common examination for all students - college-preparatory, general, and vocational. Opinions about the implications of such a system for students pursuing different curricular tracks--particularly on weaker students--were not explored.

In 1978, respondents were asked whether tests in different subjects used for promotion should be prepared at the local, state, or national level. Only 28% chose the national option, while 37% and 25% chose the local and state options respectively (Gallup, 1978). In 1987, people were asked whether they favored or opposed educational achievement test results being reported on a state-by-state and school-by-school basis so that comparisons could be made between schools of similar size and racial and economic make-up. (The source of the achievement tests is not explicit, i.e., existing commercially-available tests, NAEP, or some new national test.) Seventy percent were in favor of the proposal, only 14% were in opposition (Gallup & Clarke, 1987).

In 1980, respondents were asked which level of government should have the greatest influence in deciding what is taught in the local public schools. Sixty-eight percent chose local government, 15% state, and only 9% the federal government (Gallup, 1980). In both 1982 and again in 1986, respondents were asked to think about the future in terms of whether the federal government should have more

or less influence in determining the educational program of the local public schools. In both years, slightly over half the respondents agreed that the federal government should have less influence; about a quarter opted for more influence, while about 1 in 10 opted for the same influence. In 1986, the same sort of question was asked about the relative influence of state government. Forty-five percent chose the more influence option while 32% chose less influence and 16% opted for the same as now choice (Gallup, 1980).

The 1989 poll contained a series of questions on support for national goals, standards, curriculum, and testing (Elam & Gallup, 1989). While past polls showed strong support for local control, the results of the 1989 poll showed that a majority of respondents supported the idea of national achievement standards and goals (70%); 69% favored a standardized national curriculum; and 77% favored a standardized national test. When asked who should establish the standards, the federal and state options only drew 5% and 3% respectively; the majority (61%) chose the professional educator option while 20% chose the panel of parents and other lay people (Elam & Gallup, 1989). These data could be interpreted as showing a distrust of state and federal intervention in the standard-setting process.

However, in 1990, when asked about the President's and the fifty governors' six national goals for education for the year 2000, people strongly endorsed the goals. Nonetheless, people were profoundly skeptical about the possibility that the goals could be reached within the decade (Elam, 1990, p.42).

The data from the various Gallup polls would indicate support for a national test--even one for graduation decisions--and a national curriculum of some kind. However, before concluding that any national test or curriculum at any grade, for any purpose, would meet with public favor, questions would need to be more sharply drawn. For one thing, it would be wise to tease out the tension between local control on the one hand and a national test and curriculum on the other. Respondents were never faced with a direct confrontation between the two concepts and most likely never considered the effects of a national test or curriculum on the cherished concept of local control or even state control. It would be wise to also ask: what subjects, content, and levels would be included in a national curriculum; what the adjective "national" means (commercial, private, quasi-governmental, or governmental); what such tests would be used for, i.e., promotion (so-called gates tests), graduation, certification of the successful completion of a given level of schooling or a given type of schooling (New York State Regents), selection (college, employment), or school accountability (NAEP); what subjects/content/skills/abilities the tests would and would not measure and at what grade(s) the tests would be administered.

Further, in light of the strong sentiments in favor of local control, and that professional educators--not state or federal government officials--should set standards, questions about the sponsorship, control, operation, and funding of a national test and national curriculum need to be asked. Finally, it would be wise to investigate public opinion concerning the variety and variations of national exams in European countries.

The Evolution of Testing Policy in the United States (1950-1991)

The sheer ubiquitousness of tests at all educational levels in the United States may partly explain why public attitudes toward a national examination are so favorable. There has been an extraordinary growth of testing over the last forty years in this country. For example, from 1972 through 1985 the number of state testing programs grew from one to 34; from 1955 to 1986, the reported dollar volume of sales of tests and testing services at the elementary and secondary levels (the elementary/high school (Elhi) market) grew by almost 400%; and, examination of yearly listings of educational literature shows that from 1960 to 1988 increasingly more entries are found under the category testing than under the curriculum category (Haney, Madaus & Lyons, in press; National Commission on Testing and Public Policy, 1990; also see Quality Education for Minorities Project, 1990, for a minority perspective on the growth of testing). A national examination, therefore, may simply be seen as a logical consequence of our growing reliance on testing.

We should also note that as long ago as 1936, the Minister for Education in Prussia, then reputedly the land of examinations, observed that the number of examinations in his country was considerably less than in the United States of America, "where students are being perpetually examined" (Kandel, 1936, p. 55). This is as true today as it was half a century ago. American students at all levels, but particularly those in elementary schools, take standardized tests far more frequently than their European counterparts. However,

many proponents of a national test argue that current tests measure low-order thinking skills, do not permit meaningful national comparisons, and do not motivate students to either work hard or to strive for excellence (for example, see Educate America Inc., 1991).

What factors account for testing's growing popularity over the past forty years? Four broad social forces have interacted to create not only demand for tests, but their increasing acceptability as policy mechanisms (Haney, Madaus, & Lyons, in press). These are: (i) recurring public dissatisfaction with the quality of education and efforts to reform education; (ii) a broad shift in the importance from inputs or resources devoted to education toward outputs produced by schools (Coleman et al, 1966; Madaus, Airasian & Kellaghan, 1980); (iii) the increased bureaucratization of schooling and society (see Wise, 1979; Haney, Madaus & Lyons, in press); and (iv) an array of legislation at both federal and state levels promoting or explicitly mandating standardized-testing programs (e.g., The National Defense Education Act of 1958; The 1964 Civil Rights Act, NAEP, The Elementary and Secondary Education Act of 1965; The Education for All Handicapped Act of 1975, PL 100-297). (See Haney, Madaus & Lyons, in press, for a discussion of these and other pieces of federal legislation that have testing provisions.)

The four are by no means independent. For example, dissatisfaction with schools can lead to testing legislation, which can increase awareness of school outcomes and increase bureaucratization. Space does not permit us to elaborate on the factors except the first since it is in the reform reports which addressed dissatisfaction with education that the idea of a national examination first emerged.

Dissatisfaction and Reform: The First Wave

Since the 1950s there have been at least five major cycles of educational reform all of which have contributed greatly to the growth and acceptance of testing: (1) the Sputnik episode in the 1950s which resulted in the 1958 passage of the National Defense Education Act (NDEA) which allocated funds for local programs for testing students; (2) the Civil Rights and compensatory education movements of the 1960s which mandated various programs that incorporated a testing provision; (3) the SAT test-score decline of the 1970s which raised numerous questions about the drop in test scores over the previous decade; (4) the first wave of education-reform reports that emerged in the early 80s; and (5) more recent reform reports. We will examine the testing provisions of the more recent reform reports in more detail below. Space does not permit discussion of the first three reform phases (see Haney, Madaus & Lyons, in press, for a discussion). We shall briefly consider the first wave of reform reports which introduced the concept of national testing (for more details see Madaus, 1985); we shall then return to the testing provisions of the more recent reports.

The first wave of reform reports illustrates one of the most common uses of test results in the policy sphere, that is, to inform policy makers about the condition of education. For example, consider the now famous ominous warning of the National Commission on Excellence in Education (NCEE) that "the educational foundations of our society are presently being eroded by a rising tide of mediocrity

that threatens our very future as a Nation and a people" (National Commission on Excellence in Education, 1983, p. 5). The indicators of that mediocrity were NAEP results, the SAT-score decline, studies of functional literacy, data from the International Assessment of Educational Achievement (IEA), and test scores from the Department of Defense. Other contemporary reform reports used the same data sources to describe the state of American education.

The following recommendation in A Nation at Risk introduced the idea of a national test while at the same time making clear the distinction between national and federal auspices:

> Standardized tests of achievement (not to be confused with aptitude tests) should be administered at major transition points from one level of schooling to another and particularly from high school to college or work. The purposes of these tests would be to: (a) certify the student's credentials; (b) identify the need for remedial intervention; and (c) identify the opportunity for advanced or accelerated work. The tests should be administered as part of a nationwide (but not Federal) system of State and local standardized tests (National Commission on Excellence in Education, 1983, p. 28, emphasis added).

The Commission called for a system of certification examinations run by an agency external to the local school district. The rationale behind this recommendation is undoubtedly the desire of the NCEE to raise and maintain standards, but it may have also been heavily influenced by two of its own commissioned papers which described

European external-examination systems used to certify the successful completion of a given level of education and for admission to third level (Brod, Farnham, Mayer & McCaughey, 1982; Resnick & Resnick, 1982).

The Education Commission of the States Task Force on Education for Economic Growth recommended periodic testing of general achievement and specific skills and that promotion should be based on mastery not age. Although ambiguous, the promotion recommendation, given the context, could well be interpreted as indicating that mastery should be determined by successful test performance. The Task Force was diplomatically non-committal regarding what agency should do the recommended testing.

A more explicit recommendation for a national test can be found in High School: A report on secondary education in America. Boyer (1983) recommended a new Student Achievement and Advisement Test (SAAT) resembling the now defunct British GCE 'O'-level examinations. The SAAT presumably would be built and administered by an external agency--the College Board. The goals of the SAAT would be two-fold: first, to evaluate the academic achievement of students on the core curriculum studied; and, second, to provide advisement in decisions about work and third-level education. While the recommendation does not explicitly call for the SAAT to be used for certification or selection purposes, the potential for such a use clearly exists.

Another call for state or national external-achievement examinations came from then Secretary for Education Terence Bell. He urged states to achieve the goal of requiring all students, except

those who are not intellectually competent, to pass examinations in English, mathematics, science, social studies, and computer science thus going beyond the minimum competency testing existing in many states (Bell, 1984).

Thus, the first wave of reform reports introduced the concept of national achievement tests, although it would take another six years for the concept to gain wide acceptance and before serious attempts to initiate a national testing program would get under way.

PRESENT PROPOSALS FOR A NATIONAL EXAMINATION IN THE UNITED STATES

Over the last two years, media coverage has reflected the increasing acceptability of the idea for national standards and a national test (for example see Chira, 1989, 1991; Cohen, 1991a, 1991b; De Witt, 1991a, 1991b; Kelly, 1991; Moses, 1991; Olson, 1991; Rothman, 1990, 1991a, 1991b; Toch, 1990, 1991; U.S. News & World Report, 1991). In this section we briefly describe six current efforts and proposals for the development of a national test.

Changes in NAEP

NAEP's current governing body, The National Assessment Governing Board (NAGB), has begun the process of setting performance standards for each of three grade levels (4, 8, & 12). This process is designed to allow NAEP to report how well students are doing compared to agreed-on national standards of excellence (O'Neill, 1991). (A recent evaluation of NAGB's standard setting

process in mathematics concluded "that the achievement levels - setting process used to date is flawed and uncorrected to the point that it does not yet warrant extension to other subject areas and longer term use in mathematics; such extension should be considered only after the effects of this first release have been thoroughly evaluated." Stufflebeam, Jaeger & Scriven, 1990, p. executive summary). Further, NAGB is presently experimenting with expanding NAEP to yield state-by-state results. Currently 35 states are participating in this effort. NAEP is also developing subject frameworks which could form the basis for a national curriculum (Smith, O'Day & Cohen, 1990). These innovations were specifically prohibited by NAEP's originators.

What is currently happening to the original safeguards so vital to NAEP's birth is illuminating. First, it illustrates how attitudes toward a national test, national standards, and a national curriculum have changed dramatically over time. More importantly, current efforts to change the character of NAEP carry a clear lesson regarding the future of any national testing system. That is, testing and assessment are technologies (Madaus, 1990); they are instrumentalities put together for a purpose, to satisfy a pressing and immediate need to solve a problem to attain an end (Basalla, 1988; Ellul, 1964, 1990; Staudenmaier, 1985; Winner, 1977). Further, the history of technology shows us that "Once a process of technological development has been set in motion, it proceeds largely by its own momentum irrespective of the intentions of its originators" (Winner, 1986, p. 40). Boorstin (1978) makes a similar point, "Technology creates momentum and is irreversible. Nothing can be

uninvented...While the currents of politics and of culture can be stopped, deflected or even reversed, technology is irreversible" (p. 9; see also Schumacher, 1977). It is not inconceivable that, in the future, NAEP could be extended further to district-level comparisons or even to judging what an individual student knows or can do. Restraints on how an existing technology is used can be very tenuous indeed (Ellul, 1964, 1990). The history of testing has plenty of examples of how tests designed for one purpose have been used for another. The use of the SAT as a wall-chart indicator, as a criterion for athletic eligibility, to award merit scholarships in New York, and as a criterion for student-loan eligibility, are all examples of a national test being used for purposes other than those originally intended, despite the objections of the CEEB and ETS.

Britain offers another example of how both the purposes and locus of control of examination systems can shift over time and political parties. As we shall see, the British system evolved with a number of independent examining bodies, each with its own syllabus and set of exams. The central government was involved only through its system of inspection. However, when the present government instituted the GCSE it reorganized the more than 20 GSE and GCE boards into five examining bodies (Murphy, 1989). It has been suggested that this is a step toward establishing a single national examining board (Scrimshaw, 1987).

Public Law 100-297

Although not widely known, the country already has provision for an optional federal national test. As part of Public Law 100-297

(1988), the Augustus F.Hawkins-Robert T. Stafford Elementary and Secondary School Improvement Amendments of 1988, the 100th Congress authorized the Secretary of Education

> after consultation with appropriate State and local educational agencies and public and private organizations, to approve comprehensive tests of academic excellence or to develop such a test where commercially unavailable, to be administered to identify outstanding students who are in the eleventh grade of public and private secondary schools....The tests of academic excellence shall be tests of acquired skills and knowledge appropriate for the completion of a secondary school education (Public Law 100-297, 1988; p. 102 STAT. 247-248).

The bill also authorized the Secretary to award students who scored at a sufficiently high level a certificate signed by the Secretary within 60 days of taking the test. However, while the Congress authorized a national test and a certificate of performance based on the test, it did not appropriate funds for implementing the program. It is interesting to speculate on how the availability of such a test and the presence or lack of a certificate on the part of individual students might be used by states, school districts, businesses, and third-level institutions.

There is one provision in PL100-297 that is worthy of note in light of how European examinations operate, that is:

> The Secretary shall assure that the tests authorized by this part are conducted in a secure manner, and that the test items remain confidential so that such items may be used in future

tests (Public Law 100-297, 1988; p. 102 STAT. 248 emphasis added).

While the first clause is certainly true of European examination systems, the emphasized section is an American concept foreign to Europeans. Following administration, European external examinations--and for that matter the New York State Regents Exams--have always been readily available to students, teachers, and the general public. In fact, it is precisely this annual release that fosters the tradition of past examinations which so strongly influences instruction and learning.

The President's and Governors' National Education Goals for the Year 2000.

Soon after the promulgation of the the President and Governors' six National Education Goals for the Year 2000, it became evident that it would do little good to have goals without the capacity to monitor progress toward their achievement. The National Education Goals Panel, chaired by Governor Roy Romer of Colorado, and made up of governors, members of the Bush administration, and congressional leaders are considering the desirability of national standards and national testing. In February 1991, Governor Romer, speaking before the National Association of Secondary School Principals, opined that "We're going to have a national test, or we're going to have a national system of tests" (quoted in Report on Education Research, 1991, pp. 3-4). Similarly, the President's Education Policy Advisory Committee (PEPAC) made up of 25 educators and business people, chaired by Paul H. O'Neil, chief executive officer of the Aluminum Company of

America, have argued that national tests are the only way to reach the goals (DeWitt, 1991; Olson, 1991; Rothman, 1990). PEPAC recommended therefore that such tests be created for English, math, science, history, and geography for students in grades 4, 8, and 12. It is not yet clear whether the tests would be national-assessment tests directed at the school (i.e. sampling students across schools) or tests directed towards individual assessment.

Proposals From the Business Sector

Both The National Alliance of Business and the American Business Conference have called for the creation of a new national standardized test of basic and academic skills for high-school graduates. Reflecting the concern of the business community for more meaningful credentials, the then Secretary of Labor Elizabeth Dole created the Secretary's Commission on Achieving Needed Skills (SCANS) to develop a test of workforce skills needed by 16-year olds.

The Commission on the Skills of the American Workforce

The provocative report America's Choice: High Skills or Low Wages identifies two problems that stand in the way of a highly educated workforce: a clear standard of achievement and lack of motivation on the part of students to work hard in school. The report argues that "Other advanced industrial nations have stringent performance standards that virtually all students must meet at about age 16 and that directly affect their employment prospects" (Commission on the Skills of the American Workforce, 1990, p. 5).[4] The Commission then recommends that "A new educational

performance standard should be set for all students to meet by age 16. This standard should be established nationally and bench marked to the highest in the world" (p. 5). All students who meet this national standard of excellence would be awarded a Certificate of Initial Mastery (CIM) that would qualify them to choose among going to work, entering a college-preparatory program, or studying for a Technical and Professional Certificate (TPC).

The CIM is not intended to be a sorting mechanism like present tests but would set a "tough standard that almost everyone will reach, although not all at the same time" (p. 6). The non-sorting, tough, but achievable standard is a very important point that must never be lost sight of. However, given the proposed tripartite use of the results of the CIM, it may be difficult over time to prevent their use for sorting. Further, the impact of the tough but attainable standard for all on disinterested at-risk students will undoubtedly become a serious issue. As we shall see, a driving force in either modifying or differentiating examinations in some European systems has been precisely the unsuitability of a common examination for all students. (Also see Calder, 1990, for a description of how an examination geared to university entrance in Alberta, Canada adversely affected non-college bound students.)

The CIM would be based on performance-based examinations for which the student can explicitly prepare. The examinations in turn would be based on explicit descriptions of what students ought to know and be able to do when they leave school. Three kinds of examination, would be used: performance examinations, portfolio examinations, and project examinations (for a description of these

modalities see Commission on the Skills of the American Workforce, 1990, pp. 100-103). The envisioned system should permit students to assemble certification credentials over a period of years, perhaps beginning as early as entrance to middle school. The call for new (actually quite old)[5] assessment modalities is a widespread, interesting, and refreshing idea and is long overdue in redressing the over-reliance on the multiple-choice mode (Commission on the Skills of the American Workforce, 1990; National Commission on Testing and Public Policy, 1990; Task Force on Education of Young Adolescents, 1989;).[6] However, it is also interesting to note that one justification for creating the CIM--which can also be found in other proposals for a national test--was the success of stringent performance standards in other advanced industrialized countries. As we shall see in Europe standards are based predominantly on essay exams, with some multiple-choice and oral modalities employed, depending on the subject field not on performance portfolio and product proposal for the CIM.

The CIM proposals bear a striking resemblance to the British pupil-profiling (or record-of-achievement) schemes (Fairbairn, 1988; Murphy & Torrance, 1988). In Britain, profiling was developed as "a method of presenting information on a student's achievements, abilities, skills, experiences, and qualities from a range of assessments, and often from a range of assessors including the students themselves" (Fairbairn, 1988, p.33). Assessment information could be based on a variety of sources (observations by teachers of group work, practical work, student products, portfolios, student-performance skills, checklists, and student self-assessment)

and reported in a variety of ways (grades, marks, percentages, comments, or performance descriptors).

Three issues arise when the British experience of profiling is considered in the American context. The first relates to problems in implementation. Profiling has been found to be time-consuming and labor-intensive. Teachers are continually forced to ask questions about the scope, detail, and format of assessment records, the frequency of their compilation, and their intended audiences. In many schools, staff have difficulty in coping with sophisticated and elaborate techniques for recording information (Fairbairn, 1988). Second, proposals for profiling in Britain arose from a dissatisfaction with the way in which public external examinations dominated the work of the school. The system of profiling was designed to give teachers greater control over the assessment process and to allow the assessment of aspects of student performance which could not be assessed in a single terminal examination. Thus, the British approaches would appear to be very different from the CIM proposals which envisage profiling as part of a public-examination system. Third, when approaches to "authentic" assessment (profiling, product and performance assessment) were implemented in Britain, there was a widening of the gap in measured performance between boys and girls (in favor of girls), between socioeconomic groups, and between ethnic groups (ILEA Research & Statistics, 1990).

The Commission proposes that credentials and certification would be determined by an organization external to the schools, independent of political influence, and broadly representing educators, employers, and the citizenry at large. That organization

would set commissions in a variety of knowledge and skill areas to "establish the performance, portfolio and project examinations and oversee the professional and objective nature of the judging of these exams" (Commission on the Skills of the American Workforce, 1990, p. 104). The work of this proposed organization is perceived to be crucial to the objectivity and fairness of the CIM.

Recently, the MacArthur Foundation and the Pew Charitable Trusts awarded Mark Tucker of the National Center on Education and the Economy and Lauren Resnick of the Learning Research and Development Center at the University of Pittsburgh (NCEE/LRDC) a $2.45 million grant to develop a new national-examination system. The NCEE/LRDC initiative seems to be related to the recommendations for a national-testing system of The Commission on the Skills of the American Workforce described above; it is also germane to the SCANS proposal.

One unique feature of the NCEE/LRDC project is important to note. The plan is to create a National Board of Educational Standards that would calibrate to a common national standard exam built under the auspices of state examining boards. This aspect of the plan resembles the English model of multiple boards whose exams are monitored by a central agency and fits one of Governor Romer's predictions - a national system of tests.

Proposals of Educate America, Inc.

The most recent proposal for a national test comes from Educate America, a newly formed nonprofit, nonpartisan institute. Its

recent publication An idea whose time has come: A national achievement test for high school seniors! argues that "The education of our youth is the foundation of our nation and that foundation is crumbling. The symptoms are clear--poor performance in every conceivable academic area in comparison with our economic competitors" (Educate America Inc., 1991, p.1). As a solution, a national battery of achievement tests is proposed for all high-school seniors, public and private, designed to measure their performance in reading, writing, mathematics, American and world history, science, and geography. Such a test would, they argue, provide on-going annual assessment of progress toward our national goals and for the first time in our country's history, provide a reliable, commonly accepted indicator of accountability for students, schools, and states.

The argument for a national test is based on a number of factors. First, the achievement levels of American students are perceived to be poor. Despite the wave of reform sparked by a A nation at risk, United States children still fare badly on the SAT (verbal scores are down six points and math scores unchanged since 1986); a look at NAEP results over time is disappointing; we also fare badly in international comparisons of achievement. Educate American points out that the United States is one of the few industrialized nations "that does not have some form of a national assessment" (p. 6). (Since NAEP is a national assessment they must be referring to external exams for the certification of students.) A direct connection is made between America's lack of a national examination and the country's low international ranking. "The result of our reliance on a system without defined outcomes for students is that the United States has the

distinction of being at the bottom of most international assessments" (Educate America, Inc., 1991, p. 10, emphasis in original). However, lack of a national test seems a rather simplistic explanation for poor performance on any of the above indicators (for example, see Rotberg, 1990 for a discussion of the fallacies and dangers of international comparisons).

Second, while the President and the governors have established national goals, it is not enough and will result in little or no improvement unless there is accountability directly linked to students and the schools.

Third, current assessment methods are inadequate. NAEP is not related to a specific curriculum, there is no direct relationship to individual students or schools and the results are of little use to local educators. State competency tests are so dissimilar and given at so many different grade levels that national comparisons are not possible. College admission tests (SAT, ACT) are not directly related to the curriculum and are not taken by all students and therefore are poor indicators of educational progress. (Nonetheless Educate America used the SAT results as an indicator of educational performance.) Commercial standardized tests differ from one another in content and level of difficulty and are not administered in a secure environment.

Educate America only mentions test security in the context of administration; no mention is made of releasing the tests each year and building new ones. A cost figure of $30 per student is offered to cover the cost of their proposal. However, it is not clear whether that would cover administration and scoring or also is meant to include yearly development costs. If the latter is the case and if the European

subject-based model of exams is followed, then the $30 figure would be too low by a factor of at least three.

Fourth, a national test will motivate students to work hard. In Educate America's scheme the test would be administered in the fall of grade 12 but would not be used for graduation. Students would be motivated to do well because they would freely share the results with employers and third-level institutions. However, if a technology exists it will often be used in ways not intended by its designers. Thus, if the results are available it would be hard to stop states or school districts from eventually using them for graduation. Similarly, employers might only hire students with certain scores and colleges might insist on the results for admission, thus de facto removing the envisioned voluntary release of scores by students.

The motivational power of a national test is a common and powerful argument in current discussions about how to reform education (O'Neil, 1991). For example, consider this from Albert Shanker:

> Why do students in most other industrialized countries perform so much better than American students do? The key can be found by facing up to a harsh market reality: No system really works unless it operates with incentives, and yet our students, unlike those in most other nations, have no incentives for doing well in school... In Europe and Japan, students cannot get into college without doing well in high school and on a series of very tough exams in a variety of subjects. And their non-college-bound students have similar incentives for working hard because if they do well in school, they will be the first to get an

apprenticeship or a job. Moreover, they will be started at a higher salary than the mediocre student (Shanker, 1990, p. 21).

Present arguments about the potential motivational effects of examinations are strikingly similar to the arguments that surrounded the introduction of examinations in 19th century Britain. In proposing his Panopticon, Bentham realized that those placed in a controlled environment of surveillance (in this case, examinations) would eventually conform to the demands of the scrutineer (see also Foucault, 1979). In the context of proposals in America today, we can ask: is it envisaged that the proposed examinations will really motivate students to learn or are they being proposed solely as a "state-managed process of accommodation to the realities of economic crisis" (Hargreaves, 1988, p. 56)? The rhetoric would suggest the latter: the examination system is being proposed as a way to produce better workers which, in turn, will transform the country's educational competitiveness. There seems to be little concern for the consequences of the educational experiences which the examinations may foster for a variety of other educational goals. For example, in light of claims that competency tests for graduation may have been implicated in increasing the drop-out rate (implying negative motivation), it will be important to keep under close examination the ramifications of the motivational power of any national examination (Kreitzer, Haney & Madaus, 1989; MDC, 1988).

Japan offers a different sort of cultural example of the motivational power of external examinations. Fallows (1987) asserts that Japanese testing is not primarily about the subject-matter content or the skills measured. Rather, they are symbols or signs of a

student's effort and determination to work hard. The important inference the Japanese make from the tests is not about scholastic attainment but about effort and determination: the presence of an iron will to go through rigorous preparation.

Educate America argues that the primary way college-bound students taking the SAT or ACT prepare is by "participating in cram courses which focus on test taking skills rather than the acquisition of knowledge or higher order thinking skills" (Educate America Inc., 1991, p. 10). They argue that their proposed tests would offer all students, college-bound or not, the same opportunity to prepare by working harder during their school experience. The implication is that cramming will not be a problem. However, private "crammers" or cram courses for European examinations have existed since the 19th century (Bowler, 1983; Madaus & Kellaghan, in press) while one of the fastest growing segments of the testing industry in the United States is that of test-preparation courses and materials which teach both content and test-taking tricks aimed at beating the system (Haney, Madaus & Lyons, in press). Crammers will certainly spring up if any form of national testing that carries important consequences for students is introduced. But since these crammers are fee-paying, poorer students and those in rural areas will not have an equal opportunity to avail themselves of such preparation.

The six or seven[7] tests in the battery proposal by Educate America would measure knowledge and skills, including factual knowledge, ability to solve problems, think critically, and analyze and interpret data. The tests would reflect "state of the art assessment practices...to minimize an over-reliance on the multiple choice" (p.9).

The results would reflect what every citizen should know in order "to participate actively in a democratic society" (p. 7). This last statement sets up a very difficult validation claim that is reminiscent of the insurmountable validity problem the state of Florida had in supporting its claim that its graduation test, the Florida State Student Assessment Test, Part II (known as the functional literacy test or the "Funci Lit"), measured skills a student needed to survive as an adult (Madaus, 1983).

There is a curious stipulation in the Educate America proposal which gives rise to confusion about the nature of the tests. That is, "the tests would be developed to measure and reward students who take a greater number of courses and more difficult courses in a particular curricular area" (p. 7). Apart from the fact that, in high school, reading and writing are not "curricular areas", the stipulation could be read to mean that the tests will favor the college-bound, brighter student, unless there are, as is the case in some EC countries, different examinations in the same field (e.g., honors or pass papers, or papers set for specific courses), which would increase the cost of the examination system. Alternatively, it could mean that the distinction is built into the scoring since five levels of reporting are proposed from MASTERY WITH DISTINCTION to NOT COMPETENT. However, differentiated scoring would still place non-college bound students at a disadvantage, and it is not clear how the score itself would convey anything about an individual's curricular background.

The British experience is instructive in this regard. The intention behind the introduction of the GCSE was to have a common examination to overcome the divisive nature of separate syllabuses and

schemes of assessment for pupils expected by their teachers to reach different levels of attainment (Murphy, 1989). However, in its development, the GCSE moved away from the idea of a common exam at 16+. With national criteria came a strong emphasis on differentiation, which involved ensuring all pupils were set appropriate tasks to show what they could do rather than what they could not do, (thus, in effect, returning to a dual examination system. Further given how the grades on the GCSE were equated with the old CSE and GDE 'O'-level examinations, the GCSE is currently designed for only about 60% of students (Murphy, 1989).

The control of Educate America's proposed national testing program is vague. All students--public and private--would have to take the tests but the mandating authority is unnamed. Presumably there would be a federal mandate since it is freely admitted that a national test would be an intrusion into what has previously been the responsibility of state and local governments. Further, in the absence of a federal mandate, each state legislature or board would have to decide to join the project and then enact a mandate. Educate America envisions committees of elementary and secondary-school teachers, business people, military officials, and post-secondary educators being convened to determine the knowledge and skills measured by the test. It is not clear who would appoint these committees. Parents, state-department officials, school administrators, and representatives of advocacy groups are among many groups missing from the committees' composition.

To create an infrastructure to build the tests, administer them in a secure fashion, and score and report the results, Educate America

proposes to hire, through competitive bid, a national contractor "capable of implementing the assessment system in all 50 states" (Educate America Inc., 1991, p. 7). Given the number of students to be tested annually--presumably for security purposes on the same day--and given the proposal for state-of-the-art assessment techniques, it is doubtful that any single contractor could handle the job--a job that would be formidable even if the tests were limited to machine-scorable items. It is not clear who would select and oversee the work of the contractor. This proposal envisages a situation that differs from anything found in Europe, where the tests are developed by central examination offices (e.g., Ireland), independent, private examining boards (e.g., Britain), or state or regional examining bodies (e.g., Germany, France), and where the same test is administered on the same day at the same time.

Educate America claims that their national test would not result in a national curriculum since it would only delineate what all students should know and what skills they should possess before they complete secondary school but would not prescribe how schools should teach. This assertion is disingenuous. European schools have national curricula but do not prescribe how schools should teach. Through a tradition of past tests, however, national tests de facto constitute a curriculum and funnel teaching and learning along the fault lines of the test. Two acronyms describe what inevitably happens: WYTFIWYG--what you test for is what you get--and HYTHYT--how you test is how you teach (Madaus, 1990).[8]

Before leaving Educate America's proposal one additional point needs to be made. Educate America, located in Morristown, New

Jersey, might look to its neighboring state New York; in fact all proponents of national examinations might examine New York's experience with external state-controlled achievement testing. New York, larger than several EC countries, has since 1864, through its Board of Regents, conducted a comprehensive system of high-school achievement testing for students seeking a Regents diploma and admission to a state college (Spaulding, 1938; Tinkleman, 1966). The examinations are controlled and built by the state and in several respects are similar to European examinations. The Regents' examinations, essay and multiple-choice, test course syllabuses across a wide range of college-preparatory and vocational subjects and are administered to large numbers of students (1.5 million at grades 8 through 12 in 1990). The Regents have encountered and solved issues related to volume, cost, and scoring. Examination papers are released each year and used by students, teachers, and parents to prepare for upcoming examinations. These tests serve to motivate many students but not all and whether students graduating with a Regents diploma are better prepared for college or for work than students from five neighboring states without external examinations is a question which needs to be addressed. Such an analysis would be very helpful in deliberations about the value of a similar examination system for the nation.

It should be noted that the New York State Regents' Exams Board, like external examining boards in Europe, uses the powerful lever of the exam to change the curriculum. However, the history of examination boards also shows that they tend to be a conservative force on the curriculum. Large deviations in a new exam from the

tradition of past exams is regarded by teachers and students as unfair. Therefore, any new body, building a national exam, would need to guard against the eventual tendency to perpetuate a predictable style of teaching and learning by not violating the tradition of their past exams (Bloom, 1961; Madaus, 1985, 1988; Madaus & Kellaghan, in press; Madaus & Macnamara, 1970; Morris, 1969; Spaulding, 1938; Tyler, 1963).

THE ORIGINS AND DEVELOPMENT OF PUBLIC EXAMINATIONS IN EUROPE

The origins and development of examination systems during the 19th century in Europe are extremely complex. Three major features, however, can be detected in their development. One relates to their selective function, a second relates to the major part played by universities, and a third to the role of examinations in defining student learning--what it is students learn and how they learn. We will describe these features and then we will list some of the advantages and disadvantages that have been ascribed to public external examinations in Europe and elsewhere.

The Selective Function of Examinations

A major feature in the origins and development of public-examination systems was their selective function (see Christie & Forrest, 1981; Creswell, 1987; Goacher, 1984; Ingenkamp, 1977). This use in the university context is obvious enough. Students gained access to a university through passing the Baccalaureat in France, the

Abitur in Germany, or the relevant university board examination in England.

However, the use of examinations for selection for the civil service and professions preceded their use for university entrance. A major reason for introducing examinations in Europe in the first place was to replace the old system of patronage and nepotism for making appointments to the civil service which had secured the dominance of the aristocracy. News had been coming through from China since the 16th century by way of missionaries and travellers about the system of examinations in that country which had used written examinations since 2200 B.C. to select personnel for government positions. The Jesuits incorporated examinations into their schools (see Du Bois, 1970; Durkheim, 1979; McGucken, 1932) and, at later date, it was hoped that the introduction of examinations for selection to the civil service would ensure that the most able and talented would be recruited.

Germany (Prussia) led the way with civil service examinations and, by 1748, recruitment to offices in the judiciary and government administrative services was based on examination performance (Amano, 1990).

Following the Revolution, France adopted the German practice of using selective examinations to identify highly trained professional elites in pursuit of a powerful absolutist state. The first use of such examinations was in 1793 when an official certificate in civic virtue was required of primary-school teachers in the ecoles centrales. In 1795, a scholarship entrance examination was established for such schools (Broadfoot, 1984). In time, examining in France, though

developed at a later date than in Germany, became more frequent, covered more subject areas, and was more selective, demanding higher standards and failing more students (Amano, 1990).

In Britain, professional bodies introduced written qualifying examinations early in the 19th century - the Society of Apothecaries in 1815 and Solicitors in 1835. At a later date, other non-university examining bodies, such as the City and Guilds of London Institute and the Society of Arts (later the Royal Society of Arts) organized examinations for commercial and technical studies. Competitive examinations for admission to the Civil Service were created under the Northcote-Trevelyan scheme in 1853 and, in the same year, a scheme was designed for selection for the East India Company's Civil Service. In these several spheres, examinations followed Benthamite principles of maximizing aptitude and minimizing expense, while at the same time controlling nepotism and patronage. By 1870, almost all civil service appointments were based on examinations while all but a few branches were open to public competition (Foden, 1989; Montgomery, 1965; Roach, 1971).

Under the influence of British practice the Civil Service Act of 1883 established competitive examinations in the United States to select personnel for government service. The examinations, however, were abandoned when Congress failed to make appropriations to continue them (Du Bois, 1970).

University Influence on Examinations

In Germany, the Abitur, first introduced in 1788 as a graduation examination for the classical middle school, soon became a

qualification examination for university. Students who passed it were automatically admitted to university. The examination was used to upgrade the quality of the universities, rejecting students with poor scholastic ability (Amano, 1990).

The influence of the university can also be seen in France in the origin of the Baccalaureat, established by Napoleon in 1808, which has been traced to the 13th century determinance of the Sorbonne University (formalized by royal edict in 1598). The determinance consisted of an oral examination to decide whether students were fit to embark on the studies offered in the university (Halls, 1965). During the 19th century, higher educational institutions administered the Baccalaureat examination, which was used both to admit students to the grandes ecoles and to government service and other professions (Amano, 1990). Up to recently, the Baccalaureat granted access to a university; now for admission to the most prestigious schools further examinations have to be taken.

In Britain also, where mass public examining became a feature of education in the 19th century, the universities played an active role in the establishment of examinations. For example, Oxford and Cambridge established systems of "locals" examinations, which were marked by university "boards", to assess secondary school quality, though it was not until 1858 that the examinations were used to examine individual students. It was later again (1877) that certification was provided to students. These examinations survived up to 1918. Other universities (Dublin, Durham) followed the same path and established procedures for examining local secondary school

pupils (Lawton, 1980; Montgomery, 1965; Mortimore & Mortimore, 1984).

However, it was London, rather than the older universities (in which examining was mainly oral and in which the examination idea had never really taken strong root) that made the main contribution to the development of the external examination idea. Established as an examining body in 1836, London did not become a teaching institution until 1898. The first London matriculation examination was held in 1838, and was the earliest external school examination in Britain. The examination was conducted entirely by written papers (Kingdon & Stobart, 1988).

Although, as we have seen, competitive examinations for the civil service and for professions preceded their use in the context of university selection, the later development of examinations outside the university was strongly influenced by people who had experienced the use of internal examinations, both in secondary schools and in universities. James Booth, who was a member of the Society of Arts and is recognized as having played an important role in the development of examinations was one such person. Booth had studied at the University of Dublin where the system of examinations--both written and oral-- was more highly developed than at the other old British universities (which placed greater store on residence and meeting other requirements for graduation). The reason for the development of examinations at Dublin was at least partly because a fifth of its students were in effect 'external', attending only for examinations each term (Foden, 1989; McDowell & Webb, 1982). Booth noted and later promoted the complex procedures he had

encountered in Dublin for use at secondary-school level, leading to the establishment of a system at the Society of Arts, which was accepted as an exemplar by other systems. In an 1853 Report of the Society, it was stated that decisive testimony had been obtained in favor of 'some system of examination for provincial schools in connection with a central body, which would be empowered to grant certificates of proficiency' (quoted in Foden, 1989, p. 75).

The universities continued to control the development of the secondary-school examination system during the second half of the 19th century in Britain. While recommendations were made, for example by the Taunton Commission in 1868 and the Bryce Commission in 1895, that a central body be created which would have overall responsibility for examinations, little was done to implement them. Although the Board of Education (set up in 1899) might have seemed the appropriate body to co-ordinate, if not administer, the examination system, opposition to central control among teachers and local education authorities was widespread. Besides, the Board did not seem keen itself to assume direct control, preferring to exercise an indirect influence. Thus, no attempt was made to wrest the initiative from the universities, which were viewed favorably by the opponents of centralization (particularly, the local authorities) (Mortimore & Mortimore, 1984).

The Influence of Examination on Student Learning

The third feature of examinations--their role in defining student learning--was in evidence from the earliest days of the French Baccalaureat. Originally, the 'program' for the examination consisted

of a catalog of 500 questions from which a number of questions were randomly selected for the examination (Prost, 1968). This approach was compatible with the catechetical method of teaching, which had been popular in Europe for several centuries.

Throughout the 19th century, examinations were often developed without any institutional provision for preparing for them. Into the present century, many students prepared for university examinations, not in the university, but in other private educational institutions or in their own homes. Given this situation, it is not surprising that examinations exerted a major influence on what it was students learned. The influence continued even when formal curricula and teaching facilities were provided.

Foreshadowing contemporary claims in America of strong links between national examinations, school achievement, and increased global competitiveness, the argument was introduced from Professor Liebig of Giessen in Germany (at that time Britain's major industrial competitor and perceived to be ahead of Britain) that, "if no examination is introduced the best schemes will fail, and will produce no effect: introduce the examination, and all the rest follows of itself" (quoted in Foden, 1989, p.74). Liebig's view from Germany echoed the British view at the time that students and teachers needed to be motivated, that competition would improve motivation and learning, and that examinations were a necessary and cost-effective means of raising educational standards and securing national competitiveness.

Throughout the second half of the 19th century, examinations flourished in Britain. The growth has been attributed to the ethos of the time which was dominated by the utilitarian values and ideas of

Adam Smith and Jeremy Bentham. First, there was the conviction that self-interest was the main motive for study and that since study involved drudgery, it was necessary to reward successful work by distinction in the ways of certificates, prizes, and medals. Second, teaching was also regarded as painful drudgery. Hence, teachers needed to be motivated by pupil success or additional payment (e.g., in payment-by-results schemes). Third, it was believed that the products of learning could be measured with some exactness. This led to an emphasis on the reproduction of factual knowledge, which formed the essence of written examinations. Fourth, the state was recognized to have only limited obligations to secure the education of its citizens. When examination systems took off in the middle of the century, there was still no state-funded provision for secondary or technical education in Britain. Fifth, examinations in the educational sphere seemed to meet the important objectives of the Benthamites in industry of uniformity and standardization. Finally, examinations came to be seen as a cheap and effective method of promoting development by demonstrating the value of improved teaching and resources, thus creating incentives to local agencies to raise funds and support education and focusing attention on the need to provide more relevant curricula in schools (Foden, 1989).

As Foden (1989) has pointed out, the principles regarding method, motivation, and the effect of examinations which were embodied in the examination movement of the 19th century were almost wholly untested and unvalidated. At the time, there was little substantial critique of examinations as a technique or process and there was no serious questioning of the utilitarian values of the

examination reformers. Booth, however, seems to have anticipated some of the later challenges to examinations when in 1853, he wrote

> Objections are sometimes made to examinations, that they cannot always be depended on as true tests of proficiency--that they gave rise to crainming, and to superficial preparation. Now there is no system or plan that was ever devised that does not stand in the shade of some one or other objection. This, however, is no argument against examination as a test; it only proves that the examiner is incompetent to discharge his duties. To be a good examiner requires previous training. A well-trained examiner who knows the subject in hand, will not only gauge the knowledge but will take due note of the faculties of those who come before him. While he who confines himself to what is set down in text-books, who makes no step in deduction, who inquires into mere facts, and not into the bearing of those facts, who does not seek to look at a truth from different points of view, mistakes the duties of his office, and leaves undeveloped the powers of the instrument in his hands. An examination should not consist of strings of leading questions, nor of interrogatories to be answered by a simple yes or no. Neither should the answer be the echo of the question, nor should familiarity with mere tabulated results be sought for. An examination should be something more than the exponent of the strength of a mechanical memory. Examinations of this kind, if they do some good, do more harm. They encourage those principles of association, which rest on verbal similitude (cited in Foden, 1989, p. 78).

Booth's description of good examiners and his belief that we can build examinations which can measure what are called today higher-order thinking skills portend many of today's arguments that we need more "authentic tests". As Booth's description of good examining indicates, however, building such tests will not be easy and will be more rather than less costly and less rather than more efficient. Nor will training teachers to teach higher-order skills be easy or cheap.

Advantages and Disadvantages of External Examinations

Given the importance of external examinations for students' future careers, it is not surprising that they have attracted many apologists and critics. Among the advantages attributed to public examinations is that they are a relatively objective and impartial means of distributing educational benefits. Indeed, as we have seen, one of the reasons for their introduction in the first place was to reduce the effects of patronage and to open higher education and a range of occupations (particularly in the public service) to a wider population of students. Other advantages attributed to public examinations include the degree of national homogeneity in educational standards and practice which they engender, the sense of purpose they give to teachers' efforts, the provision of tangible incentives for students, a diminution of conflict between the roles of teaching and assessment, the provision of an assessment procedure unaffected by personal relationships between teachers and students, their acceptance in the community, and the creation of some measure of social consent among the young, while meeting some definition of comprehensiveness, equal access, and common entitlement or shared experience (Bowler, 1983;

Commission on Mathematics, 1959; Consultative Committee on Secondary Education, 1938; Curriculum and Examinations Board, 1985; Hargreaves, 1988; Heyneman, 1987; Hotyat, 1958; Madaus & Macnamara, 1970; Morris, 1969).

At about the time that James Booth was writing in the middle of the last century, criticisms of public examinations and of their effects were also beginning to appear. Since then, there has been an avalanche of observations and analyses, in Europe and elsewhere, which have catalogued the shortcomings of public-examination systems. A major criticism of such examinations is that they can have undesirable "back-wash" effects on classwork, not just in examination classes but in lower grades as well, by limiting approaches to learning. They tend to encourage undue attention to material that is covered in the examinations and, since what is examinable is limited, worthwhile educational objectives and experiences may be excluded from the classroom. In effect, the examinations may come to determine the shape of the curriculum rather than the curriculum determining the shape of the examinations and examination performance comes to be regarded by parents and students as the main, if not the sole, objective of education.

Further criticisms made of external examinations are that they are usually carried out under artificial conditions in a very limited time frame and that they are not suitable for all students and can be extremely stressful for some, causing undue strain and excessive anxiety. They are often viewed by students as unfair since doing poorly, for whatever reason, on an examination at the end of the year can over-ride a year or more of hard work and achievement. They

tend to inhibit the development of curricular variety which may be necessary to serve local and student needs. Further, there is often a lack of congruence between course objectives and examination procedures (e.g., there may be no examinations for oral or practical objectives); and, given certain kinds of teaching, it would appear that students can do well in examinations without recourse to higher levels of cognitive activity; preparation for examinations often over-emphasizes rote memorization by students and drill and practice as teaching methods. There is plenty of evidence that examinations lead to cramming.

Particular criticisms apply to essay-type examinations. Such examinations are inevitably limited in the range of characteristics which they can assess, relying heavily on verbal and logico-mathematical areas. Procedures to reduce unreliability and inconsistency in marking are time-consuming and expensive.

Finally, public examination results are often used to serve a variety of purposes for which they may not have been designed; they may be used to predict future educational and vocational performance as well as to certify the completion of a course of education, though little effort is made to match occupational or educational needs with candidates' talents (Goacher, 1984). In particular, the use of examinations for the dual purpose of certifying the completion of a secondary education and for university admission tilts the examinations towards an academic university domination. Further, examination results may be used by universities and businesses principally to control numbers and to screen out excess applicants with the cut-off score being a function of the ratio of applicants to

places. Examinations can also force students out of school before taking the examination or after failing it, or they can result in teachers excluding students who, if they sat the examination, would probably perform poorly (Amano, 1990; Bell & Grant, 1974; Bloom, 1961; Bowler, 1983; Broadfoot, 1984; Calder, 1990; Cannell, 1989; Consultative Committee on Secondary Education, 1938; Cuban, 1986; Cummings, 1980; Cunningham, 1989; Curriculum and Examinations Board, 1985; Eisemon, Patel & Abagi, 1987; Fallows, 1987; Gayen et all, 1961; Goacher, 1984; Gordon & Lawton, 1978; Haertel, 1989; Haladyna, Nolen & Hass, 1989; Holmes, 1911; Holt, 1969; Kamii, in press; Kellaghan & Greaney, in press; Kelly, 1989; Kreitzer, Haney & Madaus, 1989; Little, 1982; Madaus & Greaney, 1985; MDC, 1988; Mehrens & Kaminski, 1988; Meisels, 1989; Morris, 1969; Mukerji, 1966; Murphy, 1989; National Commission on Testing and Public Policy, 1990; Srinivasan, 1971; Rafferty, 1985; Raven, 1977; Reynolds, 1988; Rosenholtz, 1987; Shephard, 1989; Shephard & Smith, 1986; Smith & Shephard, 1988; Spaulding, 1938; Stake, McTaggart & Munski, 1985; Stodolsky, 1988; Turner, 1984; Tyler, 1963; Wheelock & Dorman, 1989; White, 1888).

The extensive literature on external examinations should serve to underline the need for serious discussion and weighing of the long and short-term cost benefits associated with the known positive and negative aspects of high-stakes testing before embarking on a national testing program.

THE STRUCTURE OF THE EDUCATIONAL SYSTEMS OF THE EUROPEAN COMMUNITY

The twelve members of the European Community are Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, and the United Kingdom. Our description of education and examination systems refer to the former Federal Republic of Germany and, in the case of the United Kingdom, to England and Wales.[9] Table 2 provides information that allows comparisons between total populations and school populations in European countries and in the United States. In the last column, American states which are equivalent in population to European countries have been identified.

The structures of an educational system and the curricula which students follow are particularly important in the context of considering examinations, since examinations are usually geared to specific curricula. Educational systems in Europe are characterized by the conventional division between primary, secondary (usually divided into lower and upper levels), and third-level education. The primary sector is relatively uncomplicated, offering a free, compulsory, and common education to all students. There have been in the past, and continue to be, differences between countries and even between states in the same country (Germany) in the length of primary schooling. The shortest primary cycle is four years (in Germany and Portugal) while the longest is seven years (in Denmark).

Second-level education has been characterized in the past by its bipartite nature (in some countries, there has been more than two parts), its selective mechanisms, and relatively low participation rates. In one type of school, a classical academic curriculum in the tradition

of the seven liberal arts (Trivium and Quadrivium) was offered. In this school type (called grammar, gymnasium, or lycee) students were prepared for third-level education and white-collar occupations. As numbers expanded, the traditional academic curriculum got watered down, subjects were presented at different levels, and practical/commercial-type subjects were introduced for some students. The type of education now offered in such schools includes university-preparation curricula but in some countries, particularly at the lower level, it also includes more comprehensive programs designed for students who are not likely to go to university. In some national descriptions, the term 'general education' is used to describe the activities of these schools.

As an alternative to more academic general education, schools offering technical curricula were established to prepare students for skilled manual occupations. In some countries, curricula were expanded in these schools, as in the case of secondary grammar schools, to accommodate the increasing numbers of students attending. By contrast with grammar schools, however, the schools provided practical, usually short-term, "continuation" education.

After the second world war, and particularly during the 1960s, demographic, social, ideological, and economic pressures led to various reviews of educational provision. The division of students at an early age was regarded as inappropriate and curricula were seen to be in need of reform. Since increasing numbers of students were staying at school to receive some secondary education, the idea of developing a system of comprehensive education at the secondary level was considered in most countries. As participation rates grew and

students remained at school beyond the lower level of secondary education, the classical academic curriculum of the upper level of secondary schools also came under pressure at the upper level, being regarded by many as inappropriate for a student body that was becoming increasingly more heterogeneous in achievement, aptitude, interest, and motivation.

Although all countries have made some moves to comprehensive lower-secondary education (up to age 15 or 16), achievements have been varied (see Wake, Marbeau & Peterson, 1979; European Communities Commission, 1987) (see Table 3). Several countries have established lower-secondary school curricula which are largely comprehensive. Denmark and Britain have gone furthest in this with ten years of comprehensive education, though there are still remnants of the bipartite system in Britain. Greece, Portugal, and Spain, which have had major reforms of education in recent years, and Italy and France have also relatively long periods of comprehensive education. The remaining countries, with the exception of Germany, can be regarded as aspiring towards, and to some extent in transition towards a more comprehensive system. However, there still exist systems which have academic-type schools, less demanding schools providing general education, and vocational/technical schools.[10]

There are some comprehensive schools in Germany but, on the whole, the German states have resisted the development of a thorough-going comprehensive system. Both major components of the traditional structure (the classical gymnasium and the vocational school) have been sufficiently strong and successful to resist possible incursions from each other. In particular, vocational education, often

seen by students as more enticing than the gymnasium-Abitur-university route, has been consolidated and improved and is generally regarded as a success of educational policy (Hearnden, 1986).
At the lower-secondary level, Germany retains four types of school: the Realschule or intermediate school (grades 5 to 10), which is the most popular and offers a technical education, the Hauptschule (grades 5 to 9), which is the least demanding, the Gymnasium (grades 5 to 13) which offers a classical academic curriculum leading to university, and the Gesamtschule or comprehensive school (grades 5 to 9/10 or 13). The situation is even more differentiated than this categorization implies as there are several types of gymnasium (classical, modern languages, scientific, economics, agriculture). The first two grades in all types of schools are conducted in accordance with comprehensive principles and are dedicated to the orientation, observation, and guidance of students. This differentiation can clearly be seen in the scope of the curriculum and how often it is revised. Since 1949 more than seven thousand separate syllabuses have been issued for general education alone. With the addition of six more states with re-unification there will be around two thousand different syllabuses in general education and at least twice that number in special and compulsory vocational education. No syllabus is valid in more than one state; no syllabus covers more than one subject; at the secondary level there are twenty different subjects; no syllabus covers more than one of the different school types or levels (Hopmann, 1991).

At the upper-secondary level, all systems offer differentiated schools and/or curricula. In the case of Italy, the system at this level

is so complicated that it has been described by Visalberghi (1985) as a 'jungle'. The system is also highly differentiated in Germany. In addition to the types of schools described above, there is a complex and changing vocational sector, in which there are several types of schools (e.g., commerce, child care, continuation, and part-time).

Table 4 provides data on the percentages of upper-secondary school students (by gender) following general education and technical/vocational curricula in countries in the European Community. In eight of the countries, a majority of students follow a curriculum of general education. However, in most countries, a sizeable number of students are enrolled in technical/vocational education courses.

In all systems, with the exception of the United Kingdom (in which the situation is changing), curricula are prescribed for schools by a central authority, usually the Ministry of Education (MOE). The definition of this prescription varies from system to system. In Germany, it is determined by each of the eleven states. In Britain, up to the present, it has, in effect, been determined at secondary level by the independent examination boards, whose syllabuses a school has chosen to adopt. In some countries (e.g., France), curricula seem to be fairly uniform across schools. In others, particularly Denmark, since great discretion regarding implementation is left to individual schools, we might expect considerable variation between schools. The trend in several countries is to allow schools a great deal of freedom in the definition of curricula during the compulsory period of schooling.

It is clear from the description which we have presented that the structures of educational systems in Europe vary considerably.

There are many other ways in which the systems differ which might have implications for the nature of examinations and for student achievement which we cannot consider here in detail. For example, there are differences in enrollment rates at various ages between countries (Table 5). While these rates in all countries are high at age 15, considerable decreases occur in some countries at ages 16 and 17. The largest decrease between ages 15 and 16 is to be found in the United Kingdom in which the participation rate drops from 99.7% to 69.3 percent. By contrast, participation in the United States at age 16 is 94.6 percent.

EXAMINATION SYSTEMS IN EUROPE TODAY

Before describing examination systems in Europe, which we will find reflect the complexity of the educational systems in which they function, two general points may be made. First, most countries have had a tradition of external examinations. And second, examination systems in all countries have been evolving as the broader social and educational contexts in which they operate have been changing. This has led in some countries to the virtual abandonment of external examinations. Following our categorization of examination systems, we will briefly consider two further issues: the idea of a single examination for all and the cost of examinations.

Tradition of External Examinations

All countries today have some elements of an external-examination system, though for some it is only at the point of entry to

third-level education. At its most general, this means that all have at some point or points in their education system more or less formalized procedures, usually separated from the classroom situation, in which a candidate has to answer questions, usually based on externally devised syllabuses, to demonstrate that he/she possesses certain knowledge and skills. Examinations most usually involve written essay-type questions, though in some countries oral examining also features. In some countries (Britain, Netherlands, Portugal, Spain) multiple-choice sections have been introduced to examinations.

On the basis of examination performance, a candidate is usually awarded a certificate or diploma which will contain information on the candidate's performance on each subject in the examination in terms of either letter grades (e.g., A, B, C, D, E), numbers (e.g., 1, 2, 3, 4, 5), or marks. Usually, grades are arrived at by simply summing marks allocated to sections of questions and across questions and papers (if an examination in a subject has more than one paper) (see Bolger & Kellaghan, 1990). The final allocation of grades may also involve some element of norm-referencing in which grade distributions in previous years are taken into account. Marks or grades (converted to points) may be summed across subjects to make decisions about university entry. European countries do little to apply to their examinations psychometric principles of the type developed in the context of testing in the United States. The two major issues of interest are objectivity and comparability. Psychometric issues such as pretesting, item-analysis, IRT, scanning, equating, reliability, and construct validity receive little or no attention. Nor does an extensive technical

psychometric community/elite exist in Europe as it does in the United States.

The certificate or diploma, in addition to testifying to the candidate's satisfactory performance in an examination in a particular subject or groups of subjects, may also confer certain rights, such as the right to be considered for (if not actually admitted) to some sector of the social, professional, or educational world. Certificates are 'credentials', analogous to stock shares or academic currency (Solberg, 1979). Thus, certification has both educational functions (testifying to and 'guaranteeing' students' standard of education) and legitimizing functions at both the level of knowledge (indicating the legitimacy of a new subject such as computer science) and of the individual (justifying the classification of individuals into social categories and the allocation of educational benefits to certain individuals).

A further feature of educational systems in Europe, which is reflected in examinations, is that syllabuses have traditionally focussed on content. At one stage, the study of classical texts was the main feature of syllabuses (Lundgren, 1986). There have been movements in several countries in curriculum development, largely under the influence of American research, to focus on "objectives", while content has broadened considerably beyond the classics (Hameyer, Frey, Haft & Kuebart, 1986). In the syllabuses for the new General Certificate of Secondary Education taken in Britain at 16+, an effort has been made to place greater emphasis on skills (Kingdon & Stobart, 1988). In the technical and trade areas also, examinations may focus on practical skills. However, syllabuses in European countries still reflect a major concern with subject-matter content and

examinations are organized in terms of traditional subject areas (languages, mathematics, sciences, history, geography, etc.). Within a subject area, syllabuses may be offered at different levels. In some countries, syllabuses for the same subject differ in different parts of the country or depending on the examination board offering the syllabus.

Finally, the tradition of examinations in Europe has been to make public the content of examinations and their results. This was so even before examinations were presented to candidates in printed form. For example, after the oral examination of candidates at Dublin University during the first part of the last century, people who had been present wrote up and circulated the questions that had been asked. These questions were then used by students who were preparing for similar examinations in the future, in effect establishing the tradition of allowing examinations to determine curricula. This became easier of course when examinations in printed form became available (McDowell & Webb, 1982). Broadfoot (1984) has noted about the present French Baccalaureat that once examination papers are published, every teacher studies them or makes their pupils try them so that the examination questions virtually become the syllabus. Indeed, in France, examinations often make front-page news in newspapers; scholars analyze questions and the cultural implications of examiners' choice of topics (Eckstein & Noah, 1989). In Germany, in addition to making examination papers available, answer scripts are returned to students who may question the way they were graded with their teachers. If a problem cannot be resolved between student and

teacher, the matter is referred to an official of the Ministry of Education.

Changes in Examination Systems

It would be misleading to think of examination systems in Europe as stable or unproblematic. Systems have always been subject to revision. However, the changes over the past three decades have been quite radical in several countries. Some countries have moved in a particular direction, only to retreat at a later date. In the mid-1970s, for example, France abolished external examinations at 16+ with the aims of postponing selection, making assessment more comprehensive, and giving a greater role to teachers in assessing students (Broadfoot, 1987b). However, the examinations were re-instituted in the 1980s, at least partly because the resources to support a school-based system of assessment had not been made available to schools. In Germany, decisions have recently been taken to attempt to restore some of the older standards which people believe have been lost by allowing candidates freedom to select subjects at lower levels of difficulty in the Abitur (Noah & Eckstein, 1990). We can expect further change as social and educational conditions alter and as those responsible for examinations respond to criticisms of systems.

A major force for change in examinations in all countries came from expanding participation rates in education which were accompanied by an increase in variance in the achievements, aptitudes, interests, and needs of students, which in turn necessitated a change in traditional curricula. This situation led many people to the

view that the formal and academic nature of traditional examinations (particularly written terminal ones), originally designed for a minority of students, was unsuitable for the assessment of many candidates and curricular areas. Criticisms of external examinations, which have been consistently made during their history, particularly ones about their narrowing effects on students' educational experiences, achieved a new level of significance and relevance when the question of submitting total populations to traditional examination procedures arose. While many commentators judged such procedures to be particularly unsuitable in the case of low-achieving students, perhaps the more striking evidence comes from students themselves who opt not to take examinations and leave school without any formal certification. In two countries with very formal external examination systems, Britain and Ireland, the number of such students (about 11% and 8% respectively) has been a matter of serious concern to policy-makers.

There have been three major approaches to changes in examination and certification systems in recent years. One has been to abolish external examinations and certification completely. The second has been to make greater use of school-based assessment for certification purposes. The third has been to shift the purpose of examinations from selection to certification and guidance.

Most European countries at some time in the past held a national school-certificate examination at the end of primary schooling. Although during their life, criticisms of these examinations were continually made, it was felt that the examinations had certain advantages, in particular that they helped to clarify for teachers the

standards that were expected, they provided a stimulus to pupils, and were useful for selection to secondary education and in seeking employment (see, for example, Ireland: Department of Education, 1954). However, in time, arguments (particularly from teachers) about the limiting effects of the examinations on the curriculum, the fact that schools prevented low-achieving students from presenting themselves for examinations by grade retention (Madaus & Greaney, 1985), changes in philosophy of education, the raising of the school-leaving age, and the provision of adequate space in secondary schools to accommodate all students led to abolition of the examinations. Of these reasons, removal of the need to select pupils for secondary education was perhaps the most compelling. No country in the European Community today operates a national external examination at the end of primary schooling. The remnants of the certificate examination exist in Italy, where school-based examinations (set, administered, and scored by pupils' own teachers) are held throughout the country and in Belgium where some schools administer an examination at the end of primary schooling. (It seems that these schools feel an examination will help to raise standards.)

While there has been a move towards greater reliance on teacher assessments at the secondary level, examinations and certification have been retained in one form or another at two points in the system in most countries: at the end of lower-secondary and at the end of upper-secondary schooling. External examinations have been abolished and certification is entirely school-based at both levels in four countries (Belgium, Greece, Portugal, Spain). Examinations in all the other countries have some input from teachers. This takes the

form either of candidates' own teachers marking examinations which have been set by an outside body, or candidates' teachers contributing assessments which are combined with the results of external examinations. While Britain from the 1960s onwards moved towards a greater dependence on candidates' teachers in the assessment procedure, this position has been reversed in recent years, during which time central government has adopted a more active role in the control of both curriculum and examinations.

A third trend in European examination systems is the shift in emphasis from selection to certification and guidance about future academic study (see, e.g., Broadfoot, 1987b). A shift in function has been possible, especially at lower educational levels, because of the expansion of places in secondary schools. Furthermore, as the numbers taking final school-leaving examinations have increased and as the examinations have become more varied, selection for traditional third-level education is no longer a concern for many students taking terminal school examinations. Increasing numbers of these are now turning to apprenticeships or technical training (Eckstein & Noah, 1989). A further possible consequence of increasing numbers and diversification of school-leaving examinations, as well as of basing school certification wholly on school-based assessment, has been the introduction of state-controlled examinations for selection to third-level education.

The selective element is evident in other aspects of examinations and indeed, despite commitment to guidance, most systems retain strong traces of their origins as instruments of selection (Eckstein & Noah, 1989). In Denmark, for example, a

student has to achieve a certain average score (5.5) in the school-based assessment at the end of the comprehensive Folkeskule at grade 10 to enter a gymnasium. In Germany, the School Leaving Certificate from the Hauptschule at grade 9 qualifies for entrance to an apprenticeship or vocational school or transfer to a Realschule. Students who leave without this certificate are awarded a non-qualifying certificate which destines them for unskilled occupations. While at the end of secondary schooling, the increase in numbers means that many students completing school are not likely to go to university and so will be interested in certification rather than in selection to a third-level institution, for those students with aspirations towards higher education, the competition in many countries (and consequently the importance of examinations for selection) has increased considerably in recent years.

A Categorization of the Systems

As we have noted, formal external examinations have disappeared at the primary-school level in all European countries. Around the end of lower-secondary education, which more or less coincides in most countries with the upper compulsory attendance age, the situation is more complex (Table 6). In six countries (Belgium, Greece, Italy, Luxembourg, Portugal, and Spain), a school-leaving certificate is awarded to students on the basis of school-based assessment (which may involve continuous assessment and/or examinations or simply testify to completion of studies). In the other six countries, examining involves a combination of internal and external procedures. The countries can be placed on a continuum in

terms of these practices from ones in which the external component of assessment is relatively small to ones in which it adopts a major role. In Germany (where there is a variety of certificates at grades 9 and 10, depending on the type of school which the student attends), examinations are set by an external agency (the MOE) but are scored by teachers in the candidates' own school. The external examination results are also supplemented by school-based assessments. In Denmark, examinations are also set by an external agency (MOE) and in addition to being scored by teachers in the candidates' own school are also scored by external teachers. In the Netherlands, the Ministry of Education sets an examination which has essay and multiple-choice components. Essay parts are marked with the aid of a marking scheme supplied to schools by students' own teachers and by teachers from another school. However, the multiple-choice component (which represents about half the written papers) is scored centrally. Oral examinations are administered by candidates' teachers. In France, examinations are set and marked by 23 regional academies, each of which produces its own examinations, based on the national curriculum. The examinations cover the subjects French, mathematics, and history/geography. For other subjects, school-based assessments are employed.

The last two countries (Ireland and the United Kingdom) show the highest level of external control in the examination system. In Ireland, the administering authority is the MOE, in the United Kingdom, five regional examining groups are responsible for the administration of the examinations. In both countries, the examinations are scored by teachers appointed by the examination

authority, who will be unaware of the identity of candidates. While there is provision in both countries, greater in the United Kingdom than in Ireland, for including the results of school-based assessment in the final examination results, the role of the candidates' own teachers is less important than in other European countries at this level.

Diversity between countries, reflecting the position at the end of lower-secondary education, is also to be found at the end of secondary schooling (Table 6). Five models are required to take account of the differences that exist between countries at the upper-secondary level. Four, as against the six countries at lower-secondary level, operate a system of certification based solely on school assessment. The four countries (Belgium, Greece, Portugal, and Spain) which fit this model at the upper-secondary level also used school-based assessment at the lower-secondary level. In the second model, the examinations are largely school-based but have an element of external control or monitoring. In Denmark and Luxembourg, the examinations are set by the MOE, but are marked by the candidates' own teachers as well as by teachers appointed by the Ministry. In Germany, the examinations are set by the MOE in each of the country's eleven states, but are marked by the candidates' own teachers. The results are sent to the MOE which identifies ones that it regards as untypical which are then discussed with the school. In the third model, which applies only in the Netherlands, the system is the same as at the lower-secondary model for that country: part of the examination is school-based and part set by the Ministry. The latter part is scored by candidates' own teachers as well as by external teachers. In the fourth model, which operates in Italy, examinations are set by the MOE and are scored by

local examination committees which include teachers from candidates' own schools. In the final model, which operates in France, Ireland, and the United Kingdom, examinations are set and scored by an outside agency--the MOE in the case of Ireland and academies and examination boards in the case of France and the United Kingdom respectively. In Ireland and the United Kingdom, there is some provision for including an element of school-based assessment.

In a number of countries, examinations beyond the school-leaving certificate examination are employed to select students for third-level education. This happens in three countries that operate a completely school-based system of assessment and certification at the secondary level (Greece, Portugal, Spain). The examinations are run by the MOE. It will be noted that these three countries have relatively low participation rates at ages 16 and 17 (Table 5) and are among the economically poorest in the European Community. In other countries, individual universities (or schools within universities) may operate their own selection system, sometimes involving testing but using other criteria as well (e.g., school record, work experience), to decide on who they will accept for a limited number of places.

One Examination for All?

A persistent theme in the American proposals is the idea that there should be one "national" examination for all students of a particular age or grade level (recommendations about the precise age or grade level vary). At this stage we can ask: how national are European examinations? Is there one examination for all? And, if not, are there problems in comparability?

As we have noted, the term "national" can mean many things. In European systems of examinations, central governments play some role. The role is most direct and influential when the MOE sets, administers, and marks examinations (as in Ireland). It is less direct and influential when the Ministry sets examinations but does not score them (as in Germany). Other types of less direct influence occur when the central Ministry has a general supervisory role in the activities of examining bodies (as in the United Kingdom). Regionalization of the administration of examinations, as occurs in France, Germany, and the United Kingdom, all large countries, dilutes the authority of the central Ministry.

Patterns of authority in examinations may vary within a country according to the type of examination and the grade level of candidates. The administration of technical and vocational examinations, which we did not consider in this paper, can be extremely complex. Sometimes such examinations are organized by the Ministry but more frequently they are carried out at the level of the school or fall within the province of specific examining groups. There are cases in which students from one country (e.g., Ireland) take examinations, particularly in the vocational technical area, but not exclusively so, which are set and marked in another country (United Kingdom). Variation relating to grade level occurs when a central body assumes authority for examining at one level, but not at others. In a number of countries, central authorities are more involved in examinations at higher levels than at lower levels (e.g., Greece, Portugal, Spain). In other countries, the same organization is responsible for examinations at all levels of the secondary sector (e.g., France, Ireland).

This description may not help very much in deciding what can be considered a "national" examination or who the authority responsible for that examination should be. Different countries have worked out different procedures in line with their school organization and traditions.

What has the European experience to say about having one examination for all students? Obviously, where examinations, even at a single level, are school-based or the responsibility of a number of authorities (regional or otherwise), there is not a common examination. Thus, in France, the area in which a student lives will determine the precise Brevet or Baccalaureat which he or she will take. In Britain, the choice for GCSE at 16+ and GCE at 18+ will depend on the choice of the school which the student attends.

Again, in countries which have differentiated secondary education, the possibility of one examination for all does not arise as long as examinations are geared to particular syllabuses which are likely to vary from one school type to another. It is impossible to get precise figures on the proportions of age cohorts who take particular examinations. The data in Table 4 on the percentages of students in general education and in technical/vocational education, taken in conjunction with enrollment rates in Table 5, may be taken as a rough indication of the proportion of students taking examinations in the two curricular areas. In Belgium, Germany, Spain, and the Netherlands, one would expect close to one half of the students still attending school not to take an academic-type examination.

It could be argued that a similar type of examination, if not the same examination, could be provided for all. Thus, we could say that a

common system of examinations rather than a common examination operates in Britain, and France at 16+, (Kingdon & Stobart, 1988). How common, however, are the experiences of examination candidates in such a system? Given the range of options available, the answer must be that great variation exists in students' experiences.

In Britain, there is a choice of examination board or boards which will usually be made by the school. Then there is the choice of subjects and subject levels--partly a school decision and partly a student one. The choice often creates problems for schools, complicated it would appear in Britain by a preference "to do worse on those examinations which carry greater prestige rather than to do better on those that are more useful" (Macintosh, 1986, p.22). Finally, the student, when he or she actually comes to sit the examination, may have a choice of questions from which to select. Thus, candidates can achieve the same grade by answering different questions (Creswell, 1987; Orr & Nuttall, 1983).

At the 18+ level, since considerable selection and attrition has taken place, the question of an examination for all does not arise. However, even if we confine our attention to those students taking examinations at the end of academic (rather than vocational) schooling in Germany and France, we find that they have a variety of options and experiences. In Germany, Abitur candidates take examinations set by different Ministries, a widely different assortment of subjects, different papers in nominally the same subject, with different weights being given to the results depending on the option chosen. The French Baccalaureat, although it retains a large core of general education subjects which all candidates are required to take, albeit

with different weights, also offers a considerable variety of options. Four options in 1950 have grown to 38 in 1988 (Noah & Eckstein, 1990).

Given these situations, it is not surprising that questions are frequently heard about the comparability of examinations operating within a single "system" (see Murphy & Torrance, 1988; Noah & Eckstein, 1990). In the context of the Baccalaureat, Eckstein & Noah (1989) have commented on the strongly demarcated hierarchy of prestige that surrounds the examination, with the mathematical options at the top and the vocational options at the bottom. They conclude that "the French examination authorities have been prepared to yield more and more comparability across candidates in an explicit trade-off to meet what are essentially political demands for 'relevance' and 'access'".

Faced with eight examining bodies at the General Certificate of Education (GCE) level and five at the General Certificate of Secondary Education (GCSE) level, the greatest efforts to achieve comparability across examinations have been made in Britain. The development of criteria for a range of individual subjects has been a step designed to help improve comparability (see Creswell, 1987; Gordon & Stobart, 1988). Further, all the examining bodies operate under the supervision of the School Examinations and Assessment Council, which, among other things, arranges for a team of examiners from one examining board to work at another board for an extended period of time, observing procedures, the marking of scripts, and standards. The Intergroup Research Committee of the boards, made up of research officers, also carries out a program of continuous research

relating to comparability between boards, subjects, and modes of examining. Staff read scripts from other boards, paying particular attention to borderline grades. To supplement these procedures, there is a strong tradition of research on British examinations (e.g., Nuttall, Backhouse & Wilmott, 1974; Nuttall & Wilmott, 1972).

It would seem from the European experience that a single common examination for all can only be provided if the content/skills assessed are relatively basic or if teachers play a major role in assessing students' performance. Attempts in Britain to provide an external examination covering a wide range of achievement do not seem to have been entirely satisfactory. While it was hoped that the General Certificate of Secondary Education would cater for a wider range of students than the GCE/CSE systems which it replaced, this does not seem to have happened. Many syllabuses have become less accessible to lower-ability candidates and in particular to those candidates who in the past would have been catered for on the basis of teacher-designed syllabuses and examinations (CSE Mode 3) (Kingdon & Stobart, 1988; Murphy, 1989). Ironically, the attempt to provide an examination with a wide-achievement span seems to have been unsuccessful not only in the case of lower-achieving students but is reported (in newspapers and on television) to have lowered the standards of the higher-achieving students who go on to do GCE A-level examinations at 18+. Teachers say that these students are less well-prepared in such areas as science and mathematics, while some university teachers have expressed the opinion that a further year at university will be necessary if students at graduation are to reach the same level as their predecessors.

The Cost of Examinations

We were not able to obtain extensive information about the costs involved in examining in European systems. In countries in which teacher assessments play a major role, costs are largely absorbed in teachers' salaries.

Information on external examinations taken at 16+, from Britain for the GCSE (in which students on average take about five subjects) and from Ireland for the Junior Certificate (in which students on average take about seven subjects), indicates that the cost of examining a student is $107. (In Ireland, candidates pay about 40% of the cost.) If the state of Massachusetts were to adopt one of these models to test its 65,000 16-year old students, the cost would be almost $7 million. At present, it spends $1.2 million to test the reading, writing, and arithmetic achievements of students at three grade levels (3, 6, 9 and 4, 8, 12 in alternate years), using machine scoring for the reading and arithmetic tests. It is clear that the adoption of an external European model of testing would have very substantial financial implications for the Commonwealth.

To test the 3 million 16-year olds in the United States would cost over $320 million using the British or Irish model. Costs would be reduced if multiple-choice tests were used, if students were examined in fewer subjects, if the range of options available to students were reduced, and if examination papers were not released. Costs would be increased very considerably if more "authentic" measures of student achievement were used. It is also likely that labor

costs (e.g. in scoring examinations) would be higher in the United States than in Britain or Ireland.

In some European countries (Britain, Ireland) and in the Canadian province of Alberta (Calder, 1990), some students repeat the last year of secondary school in an attempt to raise their already passing scores on examinations to qualify for a university place. This involves considerable expense to taxpayers which should be considered in making cost estimates of national examinations.

CONCLUSION

We may take it as axiomatic that an educational system develops in response to the values, needs, and aspirations that characterize a nation. We may also regard it as axiomatic that the components of any system interlock in an idiosyncratic way to contribute to the realization of a nation's goals for education. This is not to say that the performance of educational systems should not be reviewed from time to time to see how effectively they are performing their tasks and, in particular, to determine if they are meeting new needs that continually arise in our fast-changing society. However, it does mean that if one decides to select a feature of another educational system in the search for solutions to one's own educational problems, one should ensure in the first place that one understands the functioning of that feature in the foreign system (including any problems that have been identified in its operation) and, in the second place, that the proposed transplant is likely to be compatible with the host system.

There are many features of an educational system and of the wider socio-economic and cultural system in which education is carried out that one could nominate as being likely, in interaction with other features, to contribute to students' achievements in schools. For example, some systems (including all European countries) use inspectoral systems to monitor the quality of education in schools. Again, some systems have a longer school year than others (see Table 1), presumably providing more "time on task" for students. Systems vary in the quality of candidate they recruit to the teaching profession and in the type and length of training they provide. Most European

systems recruit better students, on average, to teacher preparation than is the case in the United States. The support which homes provide is also generally accepted as an important factor in student learning.

To isolate national examination systems from their context, reduce them to a consistent relationship with achievement, educational quality, increased global competitiveness, and suppose they have the same effect in all contexts, is based on a false sense of technological optimism. Indeed, the conviction that the establishment of national examinations, standards, and curricula will usher in a golden age of education is consistent with the great tradition of optimistic technophilia of the last two centuries (Winner, 1977, 1986). There is the further danger that focusing on examinations and assessment may lead us not only to over-look other important aspects of the system that need attention (e.g., the need to modify instructional techniques for at-risk youth) but actually to aggravate existing problems (e.g., difficult examinations and graduation requirements may lead to an increase in drop-outs) (MDC, 1988).

We do not know of any evidence that would tell us whether having or not having external examinations in the United States would have the effects which proponents of national examinations are at present suggesting. In this paper, we described the ethos in which examinations grew in Europe, the complexity of the systems in which they are embedded, and some of the problems that have been noted in their operation. Whether or not examination systems introduced to the United States would work in the same way as they do in Europe

and, in particular, whether or not they would impact on educational standards are matters that require serious consideration.

We will conclude by considering the relevance of European experience with examinations to the proposals for examining that have been made in two reform waves in the United States which we described above. Although there is considerable complexity and variation in European systems of examinations, there are, as we noted, certain features that characterize several systems. In considering American proposals, we will regard European experience as particularly relevant, if it occurs in several countries or even if it is found in only one country, we may still cite it if it seems to speak directly to an American proposal. The British experience seems to fit this latter criterion since some of the proposed American reforms seem to have been inspired by, if not actually modelled on, the British system. In applying any European experience, we should not lose sight of differences between the United States and European countries in the political, social and educational contexts in which examination systems function.

Purpose of Testing

In general terms, proposals for examinations in the United States are directed towards raising educational standards which, in turn, it is hoped will improve the economic competitiveness of the nation. It is envisaged that the examinations will have their effect by promoting a common curriculum in schools across the country and by putting pressure on teachers and students to achieve a high standard of performance on the examinations.

When one looks at proposals in more detail, however, one finds a wide variety of purposes being posited for examinations. Some of them relate to decisions within the school--the identification of students for remedial intervention or for advanced or accelerated work (National Commission on Excellence in Education) or for grade promotion (Education Commission of the States Task Force on Education for Economic Growth). Others relate to decisions which might involve leaving school. The Commission on the Skills of the American Workforce sees examinations as providing information relevant to whether a student enters a college-preparatory program, studies for a technical certificate, or goes to work. Boyer's report on secondary education regards examinations as being relevant to decisions at a later stage about whether one goes to work or to third-level education. These proposals all emphasize the use of examinations for guidance. However, they do not make clear why examinations are needed to improve the guidance services which are already in place in American schools. Neither do they face the issue of how examinations are going to motivate students to work harder if high stakes are not attached to performance on those examinations. While the proposals may not explicitly acknowledge the fact, it would seem that they envisage the use of examination results to make decisions about students, as indeed the Education Commission of the States Task Force in Education for Economic Growth does when it says that examination performance could be used in deciding grade promotion of students, and Public Law 100-297 does when it proposes the use of examinations to identify outstanding students.

Another proposed use of examinations is the certification of students' achievements. This use is proposed by the National Commission on Excellence in Education, the Commission on the Skills of the American Workforce, and Public Law 100-297. However, there is not total agreement among the reports on the use of examinations for certification. Educate America, in proposing the testing of all high-school seniors, specifies that the examinations should be held in the fall of grade 12 so that they would not be used for graduation.

Other purposes proposed for testing are accountability of students, schools, and states (Educate America) and monitoring of standards in schools (National Commission on Excellence in Education). While the idea that examinations would motivate students to work harder runs through all the proposals, it is most explicit in the Educate America proposal.

The proposals are most compatible with European practice when they emphasize the use of tests for the two inter-related functions of certification and selection. The origin and traditions of examinations in Europe exhibit a major concern with selection, particularly for third-level education and certain jobs. As the educational systems of the countries have come to resemble more that of the United States in its student retention rate and curriculum comprehensivization (up to the age of about 16), the emphasis in examinations has shifted from selection to certification. However, examination results continue to be used for selection, both inside the school system and outside it. An inevitable consequence of accepting the spirit of European examinations (even if not the details) would

seem to involve a greater emphasis on the selection and categorization of students in American schools.

A number of further aspects of the purposes of European examination systems seem relevant to the American proposals. The public-examination systems in Europe are not used formally for accountability or monitoring of standards. Rather, examinations are used to make decisions about individual students, not about teachers, schools, or districts (though parents may make judgements about schools on the basis of their examination results). Neither are examinations used to help improve or monitor standards. Efforts in Britain to allow comparability between marks from successive examinations set on the same syllabus are made so that users can be confident that a given grade has the same meaning from year to year, rather than to improve overall standards.

For the most part, quality control and accountability in European educational systems are the function of school inspectorates. All countries also participate in comparative studies of educational achievement and some have national-assessment systems similar to NAEP, in which individual school performance is not identified. While existing public-examination systems do not seem appropriate to serve the functions of monitoring or accountability, efforts are being made in Britain to develop a system which will serve these functions as well as the traditional student certification function of public examinations for students up to the age of 16 years.

While a single assessment system is unlikely to meet efficiently a variety of purposes, neither are the purposes of examinations completely independent of each other. Thus, examinations are likely

to have a motivating effect on some students only if performance on them has some real consequence for the students (e.g., admission to a third-level institution, selection for a job) or if examination performance is used for some other high-stakes purpose, such as teacher accountability.

A final point to note in considering the relevance of the purposes of examinations in Europe for the United States is that European examinations fit into a differentiated (highly so in some countries) educational structure. All systems have different types of school and curricula at the upper-secondary school level and examination performance at the end of lower secondary (about age 16) is an important factor in determining the type of school and course in which a student will find himself or herself.

Age/Grade of Testing

Apart from one proposal to test students at grades 4 and 8 (President's Education Policy Advisory Committee) and another to test at unspecified major transition points (National Commission on Excellence in Education), all the American proposals envisage examinations at either age 16 or at some point during the senior-high school years, including the point of graduation.

The proposals which confine testing to age 16+ are compatible with European practice. Formal national external examinations below the age of 15 or 16 years no longer exist in any European country. Even at age 16+, six of the twelve European community countries do not have an external examination. In two further countries, examinations are set by an external agency but are marked by

candidates' own teachers. The trend in most countries, with the exception of Britain and France during the 1980s, has been to reduce the external element in examinations at this stage (age 16+).

Britain also differs from other EC countries in that it proposes to institute national testing of students at ages 7, 11, 14, and 16, with a new form of teacher-administered tests. However, this testing, except at ages 14 and 16 will not form part of the public-examination system.

Responsibility for Testing

Proposals for national examinations in the United States are not clear about who should be assigned responsibility for the examinations. The decision, however, is an important one since the responsible agency will exercise considerable control over teaching and learning in schools. Some indications are provided in the proposals. The responsible body should be external to the school (Commission on the Skills of the American Workforce) or school district (National Commission on Excellence in Education) and it should be a national system (but not controlled by the federal government), involving state and local tests (National Commission on Excellence in Education). However, Public Law 100-297 would be a federal test under the control of the Secretary of Education. A proposal has also been made for an agent to build and administer the examinations in all states (Educate America) while another recommends the establishment of a National Board of Educational Standards to calibrate to a common national standard examinations which would be built under the auspices of state examining boards (National Center on Education and

the Economy/Learning Research and Development Center at the University of Pittsburgh).

In European countries, with the exception of Britain and Germany, the central government has a major responsibility for curricula and examinations. Up to recently in Britain, curricula were a matter for local education authorities and schools, while examinations were controlled by independent examining bodies with loose links to universities (except for the London University Examining Board which is a part of the university). In the last few years, central government has been adopting an increasingly active role in the specification of curricula and in the control of examinations. In Germany, responsibility for curricula and examinations rests with the eleven state governments.

In these two large countries, as well as in another large country, France, one authority does not assume responsibility for the administration of examinations. Such countries probably provide a better model for the United States than smaller countries in which the Ministry of Education has responsibility for public examinations.

While Ministries of Education in all European countries play a role in examinations, the important role played by other interested parties should be recognized. In the setting of papers and standards, teachers, subject specialists, and university personnel, in addition to Ministry officials (school inspectorate), play a role. Teachers in all countries also mark examinations and assign grades to candidates. At both the 16+ and 18+ levels, in nine countries, whether or not examinations are set by an authority outside the school, candidates' own teachers play a major role in marking examinations. In the other

countries, the major role in marking is played by teachers from other schools who are not aware of the identity of candidates.

Areas of Testing

Although there are references to tests of "academic excellence" (Public Law 100-297) and of what students ought to know and be able to do when they leave school, including the knowledge needed to participate in a democratic society (Commission on the Skills of American Workforce, Educate America), most of the American proposals for what should be examined are more specific. A number of reports mention basic or general achievement and skills (Education Commission of the States Task Force on Education for Economic Growth, National Alliance of Business), while several spell these out in terms of traditional subject areas. English, mathematics, science, history, and geography or social studies are proposed by the President's Education Policy Advisory Committee and by Terrence Bell, who also adds computer studies. Educate America lists the same subject areas but instead of English proposes reading and writing.

European examination systems tend to emphasize broad cultural goals in their examinations rather than preparation for later life, though the latter of course is not ignored. Some European systems provide external examinations in a small core of subjects at the 16+ level--two (Danish and mathematics) in Denmark or three (French, mathematics, History/Geography) in France. Others allow students to take a larger number of subjects and, with some restrictions, to choose the subjects they will take (Britain, Ireland). In these countries, examinations are offered at different levels. Different types

of examinations are provided in countries which have a differentiated educational structure according to the type of school attended by the student (Germany, Netherlands).

No European country offers a single examination for all students at the 18+ level. By this stage, most countries offer one system of examinations for students following academic university-oriented curricula and another system for students following more vocationally-oriented curricula. Within the academically-oriented system, students may be required to take certain core subjects in combination with options they themselves choose. A choice of levels varying in difficulty (higher/lower; honours/pass; higher/ordinary) within subjects will also be available to students. The British systems at 18+ differs from other European systems in its high level of specialization.

It is clear that systems of examination in Europe, especially during the senior high-school years, are much more complex than the systems being proposed for the United States. The European experience can contribute little to the design or implementation of proposals for a single external examination for all students at the upper-secondary school level.

Methods of Testing

Proposals about methods of testing in the American context range from the use of standardized tests (National Commission on Excellence in Education) to the use of "state of the art" assessment practices (Educate America) which, in current thinking would include the performance, portfolio, and project examinations specified by the Commission on the Skills of the American Workforce. Other proposals

recommend subject-matter examinations in core curriculum subjects as we noted above, which presumably would use the predominant European mode of having students write extended essays.

There is little that European systems of examining can tell us about the value of such procedures as portfolio and performance assessment. While efforts have been made to develop such procedures in Britain (sometimes called records of achievement), the efforts have been inspired by perceived inadequacies of public examinations to record accurately and in sufficient detail students' achievement records. This work has focused in the first place on lower-achieving "non-academic" students who were being poorly served by the examination systems, though it is hoped to extend the procedures to all students. With several competing models of assessment being developed or reformed in Britain at the moment (public examinations, Student Assessment Tasks, and profiles of achievement), it is difficult to predict what the final shape of assessment practice will be by the end of the 1990s.

One Test or One System of Testing

American proposals in some cases indicate that a single set of examinations would be used to test all students at a given age or grade level. In other cases, the suggestion is made that a system of examinations, rather than a single examination is needed.

As we have seen above, the larger European countries (Britain, France, and Germany) have a number of examination authorities that devise and administer their own tests. In France, the examinations of

the different authorities are based on a common curriculum; in Britain and Germany, they are based on separate curricula.

In two of the remaining countries that use an external examination at 16+ (Denmark, Netherlands), scripts are marked by candidates' own teachers. In the third (Ireland) students select from a range of subjects offered by the examination authority. Thus, a common examination, as distinct from a common system, exists in only two countries at 16+ (Denmark and the Netherlands).

The situation at 18+ is, as we saw, much more complex. Students' examination experience can vary, depending on the region of the country in which they live (France, Germany), the examination authority they choose (Britain), and the curriculum options they have chosen in the upper-secondary school (which vary in subject matter and in level within subject).

The question of comparability arises when students come to use their certificates either for entry to third level education or in seeking employment. In Britain and Ireland research efforts have been directed towards investigating the comparability of performance of students who take different groupings and/or levels of subjects. On the whole, however, the question of comparability of examination performance within countries does not appear to be a major problem. Rules of thumb are usually devised on the basis of a judgmental process relating to grading and comparability of grades and these are generally accepted by universities and employers.

Effects of Examinations

We have already considered in the paper many of the effects which have been attributed to examinations to which high stakes are attached. These include motivational ones ("making" teachers and students work), focusing teachers' and students' activities, cramming, emphasizing memory work, and developing test-taking skills. Here, we will just note that positive motivational effects are likely to operate only if students perceive they have a good chance of achieving the rewards attached to high test performance. For students who are not likely to do well (and thus for whom the stakes are, in effect, irrelevant) the negative effects of examinations have been a matter of serious concern in many European countries. Over the years, efforts have been made in many countries to adapt the examination system to suit these students.

It is important to note that in Europe the impact of examinations on teaching and learning--what is taught and learned and how it is taught and learned--is mediated through the availability of past examination papers. An American proposal (Public Law 100-297) not to release test papers after examinations would diminish the impact of the examinations on teaching and learning in the schools.

Cost

The only United States proposal for a national test that offers a cost estimate is that of Educate America; their figure is $30 per student. As we saw, cost figures in Europe were not readily available except for Ireland and the GCSE in Britain. As we noted, the cost for an essay on demand exam at age 16+ (consisting of between 5 and 7

separate exams in Britain and Ireland respectively) was $107. The labor costs (e.g. in scoring the exams) would probably be higher in the United States than those incurred in Britain and Ireland. In many countries the exam costs are largely absorbed in teachers' salaries. Costs would surely be higher than the $107 figure if more "authentic" assessment measures of student achievement were used. We also pointed out that repeating the last year of secondary school to improve exam scores involves considerable expense to taxpayers.

Fitting Examinations into the Existing System of Testing

With the exception of Britain, European countries do not have external systems of examinations other than the public-examination system. In most countries, little use is made of standardized tests which have been developed outside the school. The formal aspect of internal school assessment mirrors the public-examination system. Students take examinations which are similar to public examinations at the end of each school term and may take "mock" public exams some months before the actual public examinations. The United States has an extensive commercial infrastructure for developing, marketing, scoring and reporting of standardized achievement tests. Companies make their money from scoring and reporting rather than from the sale of the reusable test booklets. These tests are widely used at all levels of education.

The place of any national exam, or systems of exams within the present system of testing needs considerable thought. For example, to use the essay form (as in Europe) for the national exam while the multiple choice form continues to be widely used by states or districts

could be confusing to teachers and students. In developing a national exam, or a system of national exams an infrastructure will have to be created for developing and scoring of assessment techniques, the reporting of results and the overseeing of the entire exam operation. In Europe teachers are an integral part of the exam infrastructure as are the MOE inspectors. We would need to consider Europe's experience in this regard, particularly their trust of teachers. Discussion of the infrastructure for an American national examination system also raised serious issues of cost and quality control. In Europe, control is governmental or quasi governmental through the MOE and the established inspectorates. Cost and oversight issues associated with using commercial companies for development, scoring and reporting will need to be weighed against developing a new infrastructure for assessment in the United States.

TABLE 1

MINIMUM NUMBER OF SCHOOL DAYS EC COUNTRIES

Belgium	182
Denmark	200
Germany (North Rhine-Westphasia only)	240
Greece	175
Ireland	180
Italy	215
Luxembourg	180
Netherlands	200
Portugal	
Spain	
United Kingdom	

Source: Stichting Research voor Beleid (1988)

Table 2
Total Grade School Population for EC Countries & US

	1985 [1] Total Population	1984 [2] 6-15 1st & Lower Sec.	1984 [2] 6-11 Primary	1984 [2] 12-15 Lower, Sec.	1984 [2] 16-19 Upper Sec.	1988 [3] Comparable Total Pop in USA
Belgium	9,858,000	1,288,500	732,600	555,900	--	MI 9,240,000
Denmark	5,114,000	696,300	404,400	291,900	327,200	MO 5,141,000
France	55,170,000	8,039,300	4,578,000	3,461,300	3,426,000	CA 28,314,000 NY 17,909,000 OH 10,855,000
Germany	61,024,000	6,684,000	3,505,000	3,179,000	4,115,000	CA & NY & OH & MO
Greece	9,935,000	1,435,400	838,300	596,900	618,300	OH 10,855,000
Ireland	3,540,000	693,400	417,500	275,900	257,300	SC 3,470,000
Italy	57,141,000	8,388,000	4,773,000	3,615,000	3,806,000	CA & NY & OH
Luxembourg	367,000	42,800	24,100	18,700	21,400	WY 479,000
Netherlands	14,492,000	2,053,200	1,103,300	949,900	981,200	TX 16,841,000
Portugal	10,157,000	1,702,800	1,011,300	691,500	--	OH
Spain	38,602,000	6,511,200	3,883,800	2,627,400	2,660,400	CA & OH
U.K.	56,618,000	7,506,900	4,110,000	3,355,900	3,709,700	CA & NY & OH
USA	239,300,500	34,191,500	19,682,500	14,509,000	14,979,000	

1 = European Community Figures
2= OECD (1990) Education in OECD Counties Paris: Author
3= World Almanc & Book Facts New York: Pharos Books
Hoffman M.S. (Ed) (1987) 1987-88,

TABLE 3

DATA ON AGE OF COMPULSORY-SCHOOL ATTENDANCE AND STRUCTURE OF THE EDUCATIONAL SYSTEMS IN THE EUROPEAN COMMUNITY

	Compulsory Age Attendance	Horizontal Structure of System	Comprehensive curriculum/schools lower secondary Grades[1]	Differentiated curriculum/schools Grades
BELGIUM[2, 3]	6-16 (16-18 P-T)	6-3-3 or 6-2-2-2	7-10*	11-12
DENMARK	7-16	7-3-2 or 7-2-3	8-10	11-12
FRANCE	6-16	5-4-3	6-9	10-12
GERMANY[3]	6-15	4-6-3	5-6*	5-13
GREECE	6-15	6-3-3	7-9	10-12
IRELAND[2]	6-15	6-3-2/3	7-9*	7-12
ITALY	6-14	5-3-5	6-8	9-13
LUXEMBOURG	5-15	6-7		7-13
NETHERLANDS	6-16	6-3-3	7-10*	7-12
PORTUGAL	6-12	4-2-3-2-1	5-9	10-12
SPAIN	6-15	5-3-3(-1)	6-8	9-13
UNITED KINGDOM	5-16	6-4-2	7-10	11-12

[1] A number of countries are less advanced than others in comprehensivization of their school structures. These countries are marked with an asterisk.

[2] Belgium and Ireland have an additional two years pre-primary education integrated into the primary-school system. All other countries have provision outside the formal educational system for early childhood education.

[3] Belgium and Germany are federations. There are two states in Belgium with completely independent educational systems. There are eleven states in the former Federal Republic of Germany (16 in the new Germany). Each of the eleven states determines its curriculum under terms agreed by the Council of State Ministers of Education.

TABLE 4

PERCENTAGES OF UPPER-SECONDARY STUDENTS IN GENERAL EDUCATION AND IN TECHNICAL/VOCATIONAL EDUCATION, BY GENDER, 1985/86

	PERCENTAGE GIRLS		PERCENTAGE BOYS	
	GENERAL EDUCATION	TECHNICAL/ VOCATIONAL EDUCATION	GENERAL EDUCATION	TECHNICAL/ VOCATIONAL EDUCATION
BELGIUM [1]	56	44	53	47
DENMARK	40	60	26	74
FRANCE [2]	65 [3]	35	58 [3]	42
GERMANY [2]	51	49	57	43
GREECE	83	17	62	38
IRELAND	79	21	86	14
ITALY [4]	26	74 [5]	22	78 [5]
LUXEMBOURG	38	62	29	71
NETHERLANDS	49	51	43	57
PORTUGAL [6]	99	1	99.8	0.2
SPAIN	58	42	53	47
UNITED KINGDOM	53	47	57	43

1 Lower and upper-secondary education
2 1986/87
3 Includes upper-secondary technological education
4 1984/85
5 Includes preschool and primary teacher training
6 Technical/vocational education was abolished in 1976.
New courses were introduced on an experimental basis in 1983/84.

Source: European Communities Commission (1990), Table 3b.

TABLE 5

ENROLMENT RATES FOR AGES 15-18
IN THE EUROPEAN COMMUNITY, CANADA, JAPAN, AND U.S.A.

1987-88

	Age 15	Age 16	Age 17	Age 18
Belgium	95.8	95.5	92.7	72.0
of whom, part-time	2.2	3.6	4.6	4.6
Denmark	97.4	90.4	76.9	68.6
France	95.4	88.2	79.3	63.1
of whom, part-time	0.3	7.9	10.0	5.2
Germany [1]	100.0	94.8	81.7	67.8
of whom, part-time			0.1	
Greece [2]	82.1	76.2	55.2	43.6
Ireland [2]	95.5	83.9	66.4	39.6
Italy [3]				
Luxembourg			83.4	71.1
of whom, part-time			15.8	15.8
Netherlands [4]	98.5	93.4	79.2	59.7
of whom, part-time				
Portugal		32.1	36.9	29.2
Spain	84.2	64.7	55.9	30.4
U.K.	99.7	69.3	52.1	33.1
Canada [3]	98.3	92.4	75.7	56.9
Japan	96.6	91.7	89.3	3.2
of whom, part-time	2.6	1.9	1.7	1.4
U.S.A. [2]	98.2	94.6	89.0	60.4

1 Apprenticeship is classified as full-time education
2 1986-87
3 Excluding third-level
4 Excludes second-level part-time education

Source: OECD (1990), Table 4.2, except figures for Portugal which are for secondary education in 1983-84 and come from European Communities Commission (1990), Table 1c.

TABLE 6

EXAMINATIONS/CERTIFICATES IN EC COUNTRIES

	Primary	End of Compulsory	End of Secondary	Entrance to 3rd level
BELGIUM	Diploma Optional Kantonal school-based exams	Series of diplomas (general, technical) (school-based)	Diplomas for general, technical, professional studies (School-based)	Some university departments (engineering) requires additional test
DENMARK	None	Leaving Certificate set by MOE, marked by own teacher & external teacher.	Atrium set by MOE and marked by own teacher & external teacher. Vocational/ technical alternatives.	
FRANCE	None	Brevet de college exam set and marked by 23 academies in basic subjects & teachers' assessment in other curricular areas.	Baccalaureat set and marked by 23 academies (questions selected from centrally approved list). Three types: general, technical and vocational.	MOE Concours or exam for admission to a grand ecole (after one two years). For lower status universities Bac is sufficient.
GERMANY	None	Series of examinations/diplomas depending on type of school attended set by 11 state MOEs and marked by own teacher.	Abitur set by 11 state MOEs and marked by own teacher. Some weight to school grades	
GREECE	None	Diploma (school-based).	School-leaving diploma (school-based)	Exam set and marked by MOE & school-leaving diploma
IRELAND	None	Two external exams set and marked by MOE; some school assessment in some subjects. (To be amalgamated into one Junior Certificate in 1992).	Leaving Certificate set and marked by MOE	
ITALY	Primary Certificate under direction of MOE. Set and marked in school	Middle-school certificate and technical/ vocational qualifications. Set by MOE, marked in own school.	Exam/Diploma of General Education; Diploma of Technical Education. Set by MOE. Marked by local exam committees (including teachers form candidates' school).	Some university departments require additional exam, but usually not.

Table 6 (Cont.)

	Primary	End of Compulsory	End of Secondary	Entrance to 3rd level
LUXEMBOURG	None	School Cert of completion	Diploma de fin d'etudes Secondaires set by MOE, marked by school and outside examiners (written)	
NETHERLANDS	None	Exam/certificate based on internal assessment and national (MOE) written exam.	Exam certificate based on internal assessment and national (MOE) written exam.	
PORTUGAL	None	School cert (based on assessment by teachers).	School cert (based on assessment by teachers).	Exam set and scored by MOE
SPAIN	None	School cert (based on assessment by teachers).	Bacillerato Unificado Polivant. School cert (based on assessment by teachers).	One year later: exam set and scored by MOE (additional exams set for some university departments)
UK	None	GCSE set and marked by 5 regional boards. Incorporates some school - based assessment	General Certificate of Education, set and marked by 8 Examination Boards.	

TABLE 7

EXAMPLES OF QUESTIONS FROM THE IRISH LEAVING CERTIFICATE EXAMINATION

History (Ordinary):

Why had the Renaissance movement little direct influence in Ireland?

What was the Spanish attitude to slavery in the colonies?

Write a short paragraph on two of the following: Burgundy under Charles the Bold, The Conquest of Granada, Savonarola, The Diet of Worms, 1521 Martin Luther's writings, The Results of the Council of Trent.

History (Higher):

Write an essay on housing and farming in Ireland in Tudor times.

Why is the reign of Elizabeth I generally regarded as one of the most important periods of English history?

English (Ordinary):

Write a composition on one of the following subjects:

(a) Why I would like to be someone else.

(b) A recent magazine article ended as follows: "So there is no need to fear for Ireland's future, now that we have joined the Common Market." Give your views of this conclusion.

(c) A hero of our times.

(d) You have seen a filmed verison of a novel (or short story, or drama) that you know. Describe how the film version has affected your enjoyment of this novel (or short story, or drama).

English (Higher):

Write a composition on one of the following subjects:

(a) The tyranny of convention.

(b) "The thoughts of youth are long, long thoughts".

(c) Write an article for a serious newspaper or magazine giving your comments on the "permissive society" and on those who have allowed it to develop.

(d) Modern society is being ruined by urbanisation.

Mathematics (Ordinary):

Differentiate with respect to x:

(i) $(x^3 - 3)(x^2 - x - 4)$

(ii) $\frac{4x+1}{x^2+x+2}$

Mathematics (Higher):

Prove De Moivre's Theorem.

if $Z = \cos\theta + i \sin\theta$, prove

$x^n + Z^{-n} = 2 \cos n\theta$

and find $\sin n\theta$ in terms of Z

Prove also that

$(\sin x + i \cos x)^n = \cos n(\pi/2 - x) + i \sin n(\pi/2 - x)$, when n is a positive integer.

TABLE 7

EXAMPLES OF QUESTIONS FROM THE IRISH LEAVING CERTIFICATE EXAMINATION

History (Ordinary):

Why had the Renaissance movement little direct influence in Ireland?

What was the Spanish attitude to slavery in the colonies?

Write a short paragraph on two of the following: Burgundy under Charles the Bold, The Conquest of Granada, Savonarola, The Diet of Worms, 1521 Martin Luther's writings, The Results of the Council of Trent.

History (Higher):

Write an essay on housing and farming in Ireland in Tudor times.

Why is the reign of Elizabeth I generally regarded as one of the most important periods of English history?

English (Ordinary):

Write a composition on one of the following subjects:

(a) Why I would like to be someone else.

(b) A recent magazine article ended as follows: "So there is no need to fear for Ireland's future, now that we have joined the Common Market." Give your views of this conclusion.

(c) A hero of our times.

(d) You have seen a filmed verison of a novel (or short story, or drama) that you know. Describe how the film version has affected your enjoyment of this novel (or short story, or drama).

English (Higher):

Write a composition on one of the following subjects:

(a) The tyranny of convention.

(b) "The thoughts of youth are long, long thoughts".

(c) Write an article for a serious newspaper or magazine giving your comments on the "permissive society" and on those who have allowed it to develop.

(d) Modern society is being ruined by urbanisation.

Mathematics (Ordinary):

Differentiate with respect to x:

(i) $(x^3 - 3)(x^2 - x - 4)$

(ii) $\frac{4x+1}{x^2+x+2}$

Mathematics (Higher):

Prove De Moivre's Theorem.

if $Z = \cos\theta + i \sin\theta$, prove

$x^n + Z^{-n} = 2 \cos n\theta$

and find $\sin n\theta$ in terms of Z

Prove also that

$(\sin x + i \cos x)^n = \cos n(\pi/2 - x) + i \sin n(\pi/2 - x)$, when n is a positive integer.

FOOTNOTES

[1] The first half of the report draws heavily on material prepared for the National Commission on Testing and Public Policy.

[2] We will not deal directly with issues about a national curriculum except as they relate to proposals for a national test. For excellent descriptions of what a national curriculum might look like in the United States and how it would fit into our present system, its possibilities and difficulties, see Shanker, 1991; Smith, O'Day, & Cohen, 1990.

[3] The Exploratory Committee on Assessing the Progress of Education, then the Committee on Assessing the Progress of Education, and in 1969, the Educational Commission of the States. Since then the management and governance of NAEP has undergone various changes that are beyond the scope of this paper.

[4] There are several points in this statement that need qualification. First, the phrase 'virtually all students' may be misleading. For example, over 10% of students in England may sit for no exam at all (Great Britain Department of Education and Science, 1989). Secondly, the phrase 'stringent performance standards' also needs qualification. The factor that directly affects their employment prospects is the presence or absence of some sort of certificate of completion which in turn is based on passing or failing an examination(s). It is the distribution of educational qualifications in the population, not the educational qualifications themselves, that is the important factor. Changes in the educational requirements for entry into further education or training and occupations occur, not because of any inherent change in the nature of the further education, training or occupation, but simply because the distribution of educational qualifications in the labour force has changed over time (Kellaghan & Lewis, 1991). Performance standards as operationalized by performance on external exams may have little to do with the qualifications really need to succeed in training or on the job.

[5] Performance, product and even portfolio methods have a long history. See Madaus (1990) for a discussion of the history of various assessment modalities.

6 For example, in its 1989 report Preparing American youth for the 21st century, the Task Force on Education of Young Adolescents, points out that in the United States the assessment of student performance influences curriculum and methods of instruction and from kindergarten to graduate school teachers teach to the test. They go on to call for new assessment practices at the middle grades that closely resemble learning tasks and reflect disciplined inquiry and higher-order thinking skills among adolescents. The report also calls for the reform of state and national achievement tests (presumably commercially available achievement tests) and calls for the use of portfolios. The new assessment presumably would be school and at the most state controlled.

7 It isn't clear whether American and world history refers to one or two exams.

8 One of the authors first heard WYTFIWYG used by Hugh Burkhart of the University of Nottingham; HYTIHYT was suggested by Paul La Mahieu of the Pittsburgh public schools.

9 Information on educational and examination systems was obtained from European Communities Commission (1987), Feneville (1987), France. Ministere de l'Education Nationale (1990a, 1990b), Great Britain. Department of Education and Science (1989), Holmes (1983), Husen & Postlethwaite (1985), OECD (1982, 1985, 1986, 1990), Solberg & Meijering (1979), Witte (1986), Xochellis & Terzis (1986) as well as in personal communications from Vasco Alves, Angela Barone, Patricia Broadfoot, Mrs Francoise Connolly, Mike Creswell, Peter Hoeber, Romain Hulpia, E. Leclerq, Javier Valbuena, Leila vang Andersen, Monique Vervoort, Ernest Weis.

10 In this stage of transition, one cannot always be clear from the type of school what curricula actually are offered. For example, in Ireland which has traditional grammar schools, vocational schools, and comprehensive schools, there is no restriction on what courses the school may offer.

REFERENCES

Amano, I (1990). Education and examination in modern Japan. Translated by W.K. Cummings and F. Cummings. Tokyo: University of Tokyo Press

Anderson, B., & Pipho, C. (1984) State-mandated testing and the fate of local control. Phi Delta Kappan, 66, 209-212.

Basalla, G. (1988). The evolution of technology. New York: Cambridge University Press.

Bell, R., & Grant, N. (1974). A mythology of British education. London: Panther.

Bell, T.H. (1984). American education at the crossroads. Phi Delta Kappan, 65, 531-34.

Beymer, L. (1966). The pros and cons of the national assessment project. The Clearing House, 40, 540-543.

Bloom, B.S. (1961). Evaluation in higher education: A report of the seminars on examination reform organized by the university of New Delhi grants commission under the leadership of Dr. Benjamin Bloom. New Delhi: University Grants Commission.

Bolger, N., & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. Journal of Educational Measurement 27, 165-174.

Boorstin, D.J. (1978). The republic of technology. New York: Harper & Row.

Booth, J. (1847). Examination the province of the state - Or the outlines of a practical system for the extension of national education. London: Parker.

Bowler, R.F. (1983). Payment by results: A study in achievement accountability. Unpublished doctoral dissertation. Boston College.

Boyer, E.L. (1983). High school: A report on secondary education in America. New York: Harper & Row.

Broadfoot, P. (1984). From public examinations to profile assessment: The French experience. In P. Broadfoot (Ed.), Selection, certification and control: Social issues in educational assessment. Lewes, Sussex: Falmer.

Broadfoot, P. (1987). Social perspectives on recent developments in assessment and examination procedures in France. Irish Journal of Education, 21, 36-52.

Brod, R.I., Farnham, N., Mayer, W.V., & McCaughey, R.A. (1982). University entrance examinations and performance expectations: A report to the National Commission on Excellence in Education, 1982. Eric Document Reproduction Service No. ED 277 102.

Calder, P. (1990) Impact of diploma examinations on the teaching learning process. Edmundton Alberta: Alberta Teachers' Association.

Campbell, R.F., & Scroufe, G.R. (1965). Toward a rationale for federal-state-local relations in education. Phi Delta Kappan, 47, 2-7.

Cannell, J.J. (1989). The "Lake Wobegon" report: How public educators cheat on standardized achievement tests. Albuquerque, NM: Friends for Education.

Chira, S. (1989, December 26). Support grows for national education standards. New York Times, A1, B10-11.

Chira, S. (1991, March, 24) The big test: How to translate the table about school reform into action? New York Times, Section 4, 1&4

Christie, T. & Forrest, G.M. (1981). Defining public examiniation standards. Basingstoke, England: Macmillan

Cohen, M. (1991, January 20). National test for national goals? Boston Globe, A21.

Cohen, M. (1991, February 17). National test for seniors? Boston Globe, A16, 19.

Coleman, J.S., (1990). Foundations of social theory. Cambridge MA: Harvard University Press

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). Equality of educational opportunity. Washington, DC: Office of Education, U.S. Department of Health, Education and Welfare.

Coleman, J.S. & Husén, T. (1985). Becoming adult in a changing society. Paris: Organization for Economic Co-operation and Development

Commission on Mathematics. (1959). Program for college preparatory mathematics, USA. New York: College Entrance Examination Board.

Commission on the Skills of the American Workforce (1990). America's choice: High skills or low wages? Rochester NY: National Center on Education and the Economy.

Consultative Committee on Secondary Education (1938). Report of the Consultative Committee on Secondary Education with special reference to grammar schools and technical high schools. London: Her Majesty's Stationery Office.

Creswell, M.J. (1987). Describing examination performance: Grade criteria in public examinations. Educational Studies, 13, 247-265.

Cuban, L. (1986). Persistent instruction: Another look at constancy in the classroom. Phi Delta Kappan, 68, 7-11.

Cuban, L. (1990). Four stories about national goals for American education. Phi Delta Kappan, 72, 265-71.

Cummings, W.K. (1980). Education and equality in Japan. Princeton, NJ: Princeton University Press.

Cunningham, A.E. (1989). Eeny, meeny, miney, moe: Testing policy and priorities in early childhood education. Chestnut Hill MA: National Commission on Testing and Public Policy, Boston College.

Curriculum and Examinations Board. (1985). Assessment and certification: A consultative document. Dublin: Author.

De Witt, K. (1991a, January 14). National tests urged for public schools. New York Times, A20.

De Witt, K. (1991b, February 2). The push to consider a once taboo subject: National school tests. New York Times, 5.

Du Bois, P.H. (1970). A history of psychological testing. Boston: Allyn & Bacon

Durkheim, E. (1979). The evolution of educational thought. Translated by Peter Collins. London: Routledge & Kegan Paul.

Eckstein, M.A., & Noah, H.J. (1989). Forms and functions of secondary-school leaving examinations. Comparative Education Review, 33, 295-316.

Educate America Inc. (1991). An idea whose time has come: A national achievement test for high school seniors! Morristown, NJ: Author.

Education Commission of the States (1984). Action in the States Task Force on Education for Economic Growth Denver, CO

Eisemon, T.O., Patel, V.L., & Abagi, J. (1987). Read these instructions carefully. Examination reform and improving health education in Kenya. International Journal of Educational Development, 7, 1-12.

Elam, S.M. (1990). 22nd Annual Gallup poll of the public's attitudes toward the public schools. Phi Delta Kappan, 72, 41-53.

Elam, S.M., & Gallup, A.M. (1989). 21st Annual Gallup poll of the public's attitudes toward the public schools. Phi Delta Kappan, 71, 41-57.

Ellul, J. (1964). The technological society. New York: Vintage Books.

Ellul, J. (1990). The technological bluff. Grand Rapids, MI: Williams B. Eerdmans.

European Communities Commission (1987). The education structures in the member states of the European Communities. Brussels: Author.

European Communities Commission (1990). Girls and boys in secondary and higher education. Brussels: Author.

Fairbairn, D.J. (1988). Pupil profiling: New approaches to recording and reporting achievement. In R. Murphy, & H. Torrance, (Eds.), The changing face of educational assessment. Milton Keynes, England: Open University Press.

Fallows, J. (1987). Gradgrind's heirs. Atlantic, 259(3), 16-24.

Feneville, J. (1987). Un apercu du systeme educatif Francais. Paris: Centre International d'Etudes Pedagogique.

Finn, C.E. (1991, January 23). Reinventing local control. Education Week, 40 & 32.

Foden, F. (1989). The examiner, James Booth and the origin of common examinations. Leeds: School of Continuing Education, University of Leeds.

Foucault, M. (1979). Discipline and punish: The birth of the prison. Harmondsworth, England: Penguin.

France. Ministere de l'Education Nationale (1990a). La rentree scolaire 1990. Paris: Author.

France. Ministere de l'Education Nationale (1990b). Reperes et references statistiques sur les enseignements et la formation. Paris: Author.

Gallup, G. (1978). 10th Annual Gallup poll of the public's attitudes toward the public schools. Phi Delta Kappan, 60, 34-46.

Gallup, G. (1980). 12th Annual Gallup poll of the public's attitudes toward the public schools. Phi Delta Kappan, 62, 33-46.

Gallup, A. (1986). 18th Annual Gallup poll of the public's attitudes toward the public schools. Phi Delta Kappan, 68, 43-60.

Gallup, A.M., & Clark, D.L. (1987). The 19th annual Gallup poll of the public's attitudes towards the public schools. Phi Delta Kappan, 69, 17-30.

Gallup, A.M., & Elam, S.M. (1988). 20th Annual Gallup poll of the public's attitudes toward the public schools. Phi Delta Kappan, 70, 33-46.

Gayen, A.K., Nanda, P.D., Duari, P., Dubey, S.D., & Bhattacharyya, N. (1961). Measurement of achievement in mathematics: A statistical study of effectiveness of board and university examinations in India, Report 1. New Delhi: Ministry of Education.

Goacher, B. (1984). Selection post-16: The role of examination results. London: Methuen.

Gordon, P., & Lawton, D. (1978). Curriculum change in the 19th and 20th centuries. New York: Holmes & Meier.

Great Britain. Department of Education and Science (1989). Education statistics for the United Kingdom, 1989 edition. London: HMSO.

Haertel, E. (1989). Student achievement tests as tools of educational policy: Practices and consequences. In B. Gifford, (Ed.), Test policy and test performance: Education, language and culture. Boston: Kluwer Academic.

Haladyna, T.M., Nolen, S.B., & Hass, N.S. (1989). Report to the Arizona legislature: Test score pollution. Phoenix, AZ: Arizona State University West Campus.

Halls, W.D. (1965). Society, schools and progress in France. Oxford: Pergamon.

Hameyer, U., Frey, K., Haft, H., & Kuebart, F. (1986). Curriculum research in Europe. Lisse: Swets & Zeitlinger.

Hand, H.C. (1965). Integrity and instructional innovation. Educational Forum, 30(1), 7-16.

Hand, H.C. (1965). National assessment viewed as the camel's nose. Phi Delta Kappan, 47, 8-13.

Hand, H.C. (1966). Recipe for control by the few. Educational Forum, 30(3), 263-272.

Haney, W., Madaus, G.F., & Lyons, R. (in press). The fractured market place for standardized testing. Boston: Kluwer Academic.

Hanna, P.R. (1961). A national commission for curriculum research and evaluation. Phi Delta Kappan, 42, 331-338.

Hargreaves, A. (1988). The crisis of motivation and assessment. In A. Hargreaves, & D. Reynolds, (Eds.), Educational policy: Controversies and critiques. London: Falmer.

Hearnden, A. (1986). Comparative studies and curriculum change in the United Kingdom and the Federal Republic of Germany. Oxford Review of Education, 12, 187-194.

Heyneman, S.P. (1987). Uses of examinations in developing countries: Selection, research, and education sector management. International Journal of Educational Development, 7, 251-263.

Holmes, B. (1983). International handbook of education systems. Volume 1. Europe and Canada. New York: Wiley.

Holmes, E.G.A. (1911). What is and what might be: A study of education in general and elementary in particular. London: Constable.

Hopmann, S. (1991) The multiple realities of curriculum policy making Paper presented to the symposium on 'International perspectives on curriculum history; The social system of curriculum policy and reforms.' AERA Annual meeting Chicago, April 4, 1991.

Holt, J. (1969). The underachieving school. London: Pitman.

Hotyat, F. (1958). Evaluations in education. In UNESCO (Ed.), Reports on an International Meeting of Experts held at the UNESCO Institute for Education. Hamburg: UNESCO.

Howe, H. (1990, December 12). The gap in our thinking about kids. Education Week, 48

Husen, T., & Postlethwaite, T.N. (Eds.) (1985). The international encyclopedia of education. Research and studies. Oxford: Pergamon.

ILEA Research & Statistics (1990). Differences in examination performance. (RS 1277/80). London: Author.

Ingenkamp, K. (1977) Educational assessment. Windsor, England: NFER Publishing Co.

Ireland. Department of Education (1954). Report of the Council of Education as presented to the Minister for Education. Dublin: Stationery Office.

Kamii, C. (in press). Achievement testing in early childhood education: The games grown-ups play. Washington, DC: National Association for the Education of Young Children.

Kandel, I.L. (1936). The international examinations inquiry. Education Record, 17 (January, suppl. no. 9), 50-69.

Kellaghan, T., & Greaney, V. (in press). Using examinations to improve education: A study in fourteen African countries. Washington, DC: World Bank.

Kellaghan, T., & Lewis, M. (1991). Transition education in Irish schools. Dublin: Educational Company.

Kelly, A.V. (1989). The curriculum. Theory and practice. (3rd ed.). London: Chapman.

Kelly, D. (1991, February 5). Some educators get testy over national exam plans. USA Today, 8D.

Kerr, C. (1991, February 27). Is education really all that guilty? Education Week, 30.

Kingdon, M. & Stobart, G. (1988). GCSE examined. London: Falmer

Kreitzer, A.E., Haney, W., & Madaus, G.F. (1988). Competency testing and dropouts. In E.F. Lois Weis, and H.G. Petrie (Eds.), Dropouts from school: Issues, dilemmas, and solutions. Part II. New York: State University of New York.

Lawton, D. (1980). Politics of the school curriculum. London: Routledge & Kegan Paul.

Linn, R.L. (1983). Testing and instruction: Links and distinctions. Journal of Educational Measurement, 20, 179-189.

Little, A. (1982). The role of examinations in the promotion of the 'Paper Qualification Syndrome'. In International Labour Office (Ed.), Paper qualifications syndrome (PQS) and unemployment of school leavers: A comparative sub-regional study. Addis Ababa: International Labour Office.

Lundgren, U. (1986). European tradition of curriculum research. In V. Hameyer et. al. (Ed.), Curriculum research in Europe. Lisse: Swets & Zeitlinger.

Macintosh, H. (1986). The sacred cow of coursework. In C. Gipps (Ed.), The GCSE: An uncommon examination. London: University of London Institute of Education.

Madaus, G.F. (1983). Minimum competency testing for certification: The evolution and evaluation of test validity. In G.F. Madaus, (Ed.), The courts, validity and minimum competency testing. Boston: Kluwer-Nijhoff.

Madaus, G.F. (1985). Public policy and the testing profession - You've never had it so good? Educational Measurement Issues and Practice, 4(4), 5-11.

Madaus, G.F. (1988). The influence of testing on the curriculum. In L. Tanner, (Ed.), Critical issues in curriculum. Eighty-seventh Yearbook of the National Society for the Study of Education, Part 1. Chicago: University of Chicago Press.

Madaus, G.F. (1990). Testing as a social technology: The inaugural Boisi lecture in education and public policy. Center for the Study of Testing Evaluation and Public Policy, Boston College.

Madaus, G.F., Airasian, P.W., & Kellaghan, T. (1980). School effectiveness: A reassessment of the evidence. New York: McGraw Hill.

Madaus, G.F., & Greaney, V. (1985). The Irish experience in competency testing. Implications for American education. American Journal of Education, 93, 268-294.

Madaus, G.F., & Kellaghan, T. (in press). Curriculum evaluation and assessment. In P.W. Jackson, (Ed.), Handbook of research on curriculum. New York: Macmillan.

Madaus, G.F., & Macnamara, J. (1970). Public examinations: A study of the Irish leaving certificate. Dublin: Educational Research Centre, St Patrick's College.

McDowell, R.B., & Webb, D.A. (1982). Trinity College Dublin 1592-1952: An academic history. Cambridge: Cambridge University Press.

McGucken, W.J. (1932). The Jesuits and education. Milwaukee WI: Bruce.

MDC (1988). America's shame, America's hope. New York: Charles Stewart Mott Foundation.

Mehrens, W.A., & Kaminski, J. (1988). Using commercial test preparation materials: Fruitful, fruitless, or fraudulent? Washington DC: National Council on Measurement in Education.

Meisels, S.J. (1989). High-stakes testing in kindergarten. Journal of Educational Leadership, 46(7), 16-22.

Montgomery, R.J. (1965). Examinations: An account of their evolution as administrative devices in England. London: Longman.

Morris, G.C. (1969). Educational objectives of higher secondary school science. University of Sydney, Australia.

Mortimore, J., & Mortimore, P. (1984). Secondary school examinations. London: University of London Institute of Education.

Moses, S. (1991, January). Motivation neglected in educational reform. APA Monitor, 34.

Moses, S. (1991, February). 'Educational excellence' goal is unmet. Monitor, 40.

Mukerji, S.N. (1966). History of education in India: Modern period. Baroda: Acharya Book Depot.

Murphy, R. (1989). The birth of the GCSE examination. In A. Hargreaves, & D. Reynolds, (Eds.), Educational policy: Controversies and critiques. London: Falmer.

Murphy R., & Torrance, H. (Ed.) (1988). The changing face of educational assessment. Milton Keynes, England: Open University Press.

National Center on Education and the Economy (1989). To secure our future: The federal role in education. Rochester, NY: Author.

National Commission on Excellence in Education (1983). A nation at risk: The imperative for educational reform. Washington DC: US Government Printing Office.

National Commission on Testing and Public Policy (1990). From gatekeeper to gateway: transforming testing in America. Chestnut Hill, MA: National Commission on Testing and Public Policy, Boston College.

Noah, H.J., & Eckstein, M.A. (1990). Trade-offs in examination policies: An international comparative perspective. In P. Broadfoot, R. Murphy, & H. Torrance (Eds.), Changing educational assessment: International perspectives and trends. London: Routledge.

Nuttall, D.L., Backhouse, J.K., & Willmott, A.S. (1974) Comparability of standards between subjects. London: Evans/Methuen Educational.

Nuttall, D.L., & Willmott, S. (1972). British examinations Techniques of analysis. Windsor, England: NFER.

O'Neill, J. (1991). Drive for national standards picking up steam. Educational Leadership, 48(5), 4-9.

OECD (1982). Reviews of national policies for education: Greece. Paris: OECD.

OECD (1985). Reviews of national policies for education: Educational reforms in Italy. Paris: OECD.

OECD (1986). Reviews of national policies for education: Spain. Paris: OECD.

OECD (1990). Education in OECD countries 1987-88. Paris: Author.

Olson, L. (1991, January 23). Advisory panel presents national-test plan to Bush. Education Week, 25.

Orr, L. & Nuttall, D.L. (1983). Determining standards in the proposed single system of examining at 16+. London: Schools Council

Prost, A. (1968). Histoire de l'enseignement en France 1800-1967. Paris: Colin.

Public Law 100-297 (1988). Augustus F. Hawkins-Robert T. Stafford Elementary and Secondary School Improvement Amendments of 1988. Public Law 100-297. 102 Stat. 130.

Quality Education for Minorities Project (1990). Education that works: An action plan for the education of minorities. Cambridge, MA: MIT Press.

Rafferty, M. (1985). Examinations in literature. Perceptions from non technical writers of England and Ireland from 1850 to 1984. Unpublished doctoral dissertation, Boston College.

Raven, J. (1977). Education, values and society: The objectives of education and the nature and development of competence. London: Lewis.

Report on Education Research (1991, March 20) Principals give a thumbs down to a national test. Alexandria, VA; Capitol Publications, Inc.

Resnick, D.P. & Resnick, L.B. (1982). Standards, curriculum, and performance: A historical and comparative perspective: A report to the National Commission on Excellence in Education, 1982. Eric Document Reproduction Service No. ED 277 104.

Reynolds, D. (1988). Better schools? Present and potential policies about the goals, organization, and management of secondary schools. In A. Hargreaves, & D. Reynolds, (Eds.), Educational policy: Controversies and critiques. London: Falmer.

Roach, J. (1971). Public examinations in England 1850-1900. Cambridge: University Press.

Rosenholtz, S.J. (1987). Education reform strategies: Will they increase teacher competence? Canadian Journal of Education, 95, 534-562.

Rotberg, I.C. (1990). I never promised you first place. Phi Delta Kappan, 72, 296-300.

Rothman, R. (1990). 2 groups laying plans to develop national exams. Education Week, 1 & 14.

Rothman, R. (1991a, February 6). Group unveils plan for national test for all high-school seniors. Education Week, 5.

Rothman, R. (1991b, January 30). Promise, pitfalls seen in creating national exams. Education Week, 1 & 17.

Saylor, G. (1970). National assessment: Pro and con. The Record, 71(4), 588-597.

Schumacher, E.F. (1977). A guide for the perplexed. London: Abacus.

Scrimshaw, P. (1987). Towards a conservative curriculum. In T. Horton, (Ed.), GCSE: Examining the new system. London: Harper & Row.

Shanker, A. (1990). Perestroika for the classroom. New Perspectives Quarterly, 7(4), 20-22.

Shanker, A. (1991, February 24). An American revolution in education: Developing a common curriculum. New York Times, E-7.

Shepard, L.A. (1989). Inflated test score gains: Is it old norms or teaching to the test. Paper presented at Symposium on 'Cannell revisited: Accountability, test score gains, normative comparisions and achievement.' AERA, 1989, San Francisco.

Shepard, L.A., & Smith, M.L. (1986). Synthesis of research on school readiness and kindergarten retention. Educational Leadership, 44(3), 78-86.

Shepard, L.A., & Smith, M.L. (1989). Escalating academic demand in kindergarten: Counterproductive policies. Elementary School Journal, 89, 135-146.

Smith, M.L., & Shepard, L.A. (1988). Kindergarten readiness and retention: A qualitative study of teachers' beliefs and practices. American Educational Research Journal, 25, 307-333.

Smith, M.S., O'Day, J., & Cohen, D.K. (1990). National curriculum, American style: Can it be done? What might it look like? American Educator, (Winter), 10-17, 40-47.

Solberg, W. (1979). School leaving examinations. Why or why not? In F.M. Ottobre (Ed.), Criteria for awarding school leaving certificates. Oxford: Pergamon.

Solberg, W., & Meijering, P.H. (1979). School leaving examinations in the Netherlands. In F.M. Ottobre (Ed.), Criteria for awarding school leaving certificates. Oxford: Pergamon.

Spaulding, F.T. (1938). High school and life: The Regents inquiry into the character and cost of public education. New York: McGraw Hill.

Srinivasan, J.T. (1971). Annual terminal examination in the Jesuit high schools of Madras, India. Unpublished doctoral dissertation, Boston College.

Stake, R.E., McTaggart, R., & Munski, M. (1985). An Illinois pair: A case study of school art in Champaign and Decatur. Urbana IL: CIRCE, University of Illinois.

Staudenmaier, J.M. (1985). Technology's storytellers: Reweaving the human fabric. Cambridge, MA: MIT Press.

Stodolsky, S.S. (1988). The subject matters: Classroom activity in math and social studies. Chicago: University of Chicago Press.

Stichting Research voor Beleid (1988). The conditions of service of teachers within the member states of the European Community. Leiden: Author.

Stufflebeam, D, Jaeger, R. and Scriven, M. (1990) Second interim assessment of NAGB's process of setting achievement levels on the 1990. National Assessment of Educational Progress in mathematics. Kalamazoo, MI; The Evaluation Center, Western Michigan University

Task Force on Education of Young Adolescents (1989). Turning points: Preparing American youth for the 21st century. New York: Carnegie Corporation of New York.

Tinkelman, S.N. (1966). Regents examinations in New York State after 100 years. In Proceedings of the 1965 unvitational conference on testing problems. Princeton, NJ: Educational Testing Service.

Toch, T. (1990, December 31). Educators embrace national exam. US News and World Report, 56.

Toch, T. (1991, February 18). Going back to school. US News and World Report, 55-56.

Turner, G. (1984). Assessment in the comprehensive school: What criteria count? In P. Broadfoot, (Ed.), Selection, certification, and control. New York: Falmer.

Tyler, R.W. (1963). The impact of external testing programs. In W.G. Findley, (Ed.), The impact and improvement of school testing programs. Sixty-second Yearbook of the National Society for the Study of Education. Chicago: NSSE.

Tyler, R.W. (1965). Assessing the progress of education. Phi Delta Kappan, 47, 13-16.

Visalberghi, A. (1985). Italy: System of education. In T. Husen & T.N. Postlethwaite (Eds.), The international encyclopedia of education. Oxford: Pergamon.

Wake, R.A., Marbeau, V., & Peterson, A.D.C. (1979). Innovation in secondary education in Europe. Strasbourg: Council of Europe.

Wheelock, A., & Dorman, G. (1989). Before it's too late: Dropout prevention in the middle grades. Boston, MA: Massachusetts Advocacy Center.

White, E.E. (1888). Examinations and promotions. Education, 8, 519-522.

Winner, L. (1977). Autonomous technology: Technic-out-of-control as a theme in political thought. Cambridge, MA: MIT Press.

Winner, L. (1986). The whale and the reactor: A search for limits in an age of high technology. Chicago: University of Chicago Press.

Wise, A.E. (1979). Legislated learning: The bureaucratization of the American classroom. Berkeley, CA: University of California Press.

Witte, G.B. (1986). Curriculum research in Spain. In U. Hameyer et al (Eds.), Curriculum research in Europe. Lisse: Swets & Zeitlinger.

Xochellis, P., & Terzis, N. (1986). Curriculum research in Greece. In U. Hameyer et al (Eds.), Curriculum research in Europe. Lisse: Swets & Zeitlinger.