

CHAPTER 5

How Other Countries Test

Highlights

- There are fundamental differences in the history, purposes, and organization of schooling between the United States and other industrialized nations. Comparisons between testing in the United States and in other countries should be made prudently.
- The primary purpose of testing in Europe and Asia is to control the flow of young people into a limited number of places on the educational pyramid. Although many countries have recently implemented reforms designed to make schooling available to greater proportions of their populations, testing has remained a powerful gateway to future opportunity.
- No country that OTA studied has a single, centrally administered test used for the multiple functions of testing.
- Standardized national examinations before age 16 **have all** but disappeared from Europe and Asia. The United States is unique in its extensive use of examinations for young children.
- Only Japan uses multiple-choice tests as extensively as the United States. In most European countries, students are required to write essays “on demand.”
- Standardized tests in other countries are much more closely tied to school syllabi and curricula than in the United States.
- Commercial test publishers play a much more influential role in the United States than in any other country. In Europe and Asia, tests are usually established, administered, and scored by ministries of education.
- Testing policies in almost every industrialized country are in flux. The form, content, and style of examinations vary widely across nations, and have changed in recent years.
- Teachers have considerably greater responsibility for development, administration, and scoring of tests in Europe and Asia than in the United States.

International comparisons of student test scores have become central to the debate over reform of American education. Reports suggesting that American students rank relatively low compared to their European and Asian peers, especially in mathematics and science, have coincided with growing fears of permanent erosion in America’s economic competitiveness, and have become powerful weapons in the hands of school reformers of nearly every ideological stripe.

A recent addition to this arsenal of comparative education politics is the examination system itself: many education policy analysts in the United States who envy the academic performance of students in Europe and Asia also envy the structure, content, and

administration of the examinations those children take. In the current debate over U.S. testing reform options, it is common to hear rhetoric about the advantages of national examinations in other industrialized countries; some commentators have gone so far as to suggest that tougher examinations in the United States, modeled after those in other countries, could motivate greater diligence among students and teachers and alter our slipping global competitiveness.²

But these arguments are based on an exaggerated sense of the role of schools in explaining broad economic conditions, and on misplaced optimism about the effects of more difficult tests on improving

¹Material in this chapter draws extensively on the OTA contractor report by George F. Madaus, Boston College, and Thomas Kellaghan, St. Patricks College, Dublin, “Examination Systems in the European Community: Implications for a National Examination System in the United States,” April 1991.

²See, e.g., Robert Samuelson, “The School Reform Fraud,”

Post, June 19, 1991, p. A19.

education.³ The rhetoric that advocates national testing using the European model tends to neglect differences in the history and cultures of European and Asian countries, the complexities of their respective testing systems, and the fact that their education and testing policies have changed significantly in recent years.

Explaining international differences in test scores is a delicate business.⁴ Similarly, drawing inferences from other countries' testing policies requires attention to the educational and social environments in which those tests operate. As a backdrop to the analysis in this chapter, it is important to keep in mind some basic issues affecting the usefulness of international comparisons of examination practices.

- Testing policies are in transition in most industrialized countries, where the pressures of a changing global economy have a ripple effect on public perceptions of the adequacy of schooling.
- Parents in Europe and Asia, like their counterparts in the United States, tend to praise their own children's schools while decrying the decline in standards and quality overall.⁵
- There is considerable variation in the structures and conduct of school systems within Europe and Asia. For example, there is probably as much difference in the degree of centralization of curriculum between Germany and France as there is between France and the United States. These differences are reflected in testing policies that vary from country to country in important ways. In Australia, Germany, Canada, or Switzerland, for example, provincial (or

State) governments have considerably more autonomy in the design and administration of tests than in France, Italy, Sweden, or Israel. Test format differs too: Japan relies heavily on multiple choice and Germany still uses oral examinations, while in most other countries the dominant form is "essay on demand."

- The functions of testing have different historical roots in Europe and Asia than in the United States. Steeped in the traditions of Thomas Jefferson, Horace Mann, and John Dewey, the American school system has been viewed as the public thoroughfare on which all children journey toward productive adulthood. Universal access came relatively later in Europe and Asia, where opportunities for schooling have traditionally been rationed more selectively and where the benefits of schooling have been bestowed on a smaller proportion of the population. Although recent reforms in many European countries have opened doors to greater proportions of children, the role of tests has remained principally one of "gatekeeper"—especially at the transition from high school to postsecondary.⁶ In this country higher education is available to a greater proportion of college-age children than in any other industrialized country.
- There is considerable variation among European and Asian countries with respect to both the age at which key decisions are made and the permanence of those decisions. For example, second chances are more likely in the United States and Sweden than in most other countries, which do not provide many options for students

³See, e.g., Clark Kerr, "Is Education Really All That Guilty?"

(New York, NY: Harper and Row, 1990); and Richard Murnane, "Education and the Productivity of the Work Force: Looking Ahead," *Institution*, 1988), pp. 215-246.

⁴See Iris Rotberg, "I Never Promised You First Place," Edward Haertel, John Schwillie, and Judith Torney-Purta,

postsecondary education ought to be factored into international comparisons, see Michael Kirst, "The Need to Broaden Our Perspectives Concerning America's Educational Attainment,"

⁵James Irving, director of Learning and Assessment Policy Division, New Zealand Ministry of Education, personal communication, February 1990.

For the United States, the latest Gallup poll shows ratings of public schools have remained basically stable since 1984. The most striking aspects are the higher ratings the public in general give their local schools (42 percent rate them an "A" or "B") compared to the grades they give the Nation's schools overall (only 21 percent rate them an "A" or "B"). Most significant, however, is the enormous confidence parents of children currently in school give to the schools their own children attend (73 percent rate these schools an "A" or "B"). It is suggested that the more firsthand knowledge one has about the public schools, the more favorable one's perception of them. Stanley M. Elam, Lowell C. Rose, and Alic M. Gallup, "The 23rd Annual Gallup Poll of the Public's Attitudes Toward the Public Schools,"

⁶See Max A. Eckstein and Harold J. Noah, "Forms and Functions of Secondary-School Leaving Examinations,"

vol. 33, No. 3, August 1989, p. 303. It is important to note that Japanese children enjoy considerably greater access to schooling than is commonly believed. For a summary of myths and data regarding Japanese education, see William Cummings, "The American Perception of Japanese Education,"

vol. 25, No. 3, September 1989, pp. 293-302.

vol. 10, No. 3, Feb. 27, 1991, p. 30; Lawrence Cremin, *Popular*

Standards, Robert E. Litan, Robert Z. Lawrence, and Charles L. Schultze (eds.) (Washington, DC: Brookings

Delta Kappan, vol. 72, No. 4, December 1990; and the rejoinder by Norman Bradburn, *Delta Kappan*, vol. 10, June 1991, pp. 774-777. For discussion of how American

who bloom late or have not done well on tests. In Japan, children are put on a track early on: the right junior high school leads to the right high school, which leads to the right university, which is the prerequisite for the best jobs. Japanese employment reflects the rigidity that begins with schooling: job mobility is negligible, “career-switching a totally alien concept. Employment opportunities for French, German, and British students are significantly affected, albeit in varying degrees, by performance on examinations.

The purpose of this chapter is to consider lessons for U.S. testing policy that can be drawn from the experiences of selected European and Asian countries. The first section provides an overview of education and testing systems in the European Community (EC) and other selected countries. The second considers lessons for U.S. testing policy. The last section contains “snapshots” of examination systems in selected countries.

Teaching and Testing in the EC and Other Selected Countries⁷

Origins and Purpose of Examinations

The university has always played a central role in examination systems in most European countries.⁸ In France, for example, the *Baccalaureat* (or *Bac*) was established by Napoleon in 1808 and has been traced to the 13th century *determinance*, an oral examination required for admission to the Sorbonne. The *Bac* was the passport to university entrance in France until recently, when additional admissions requirements were developed by the more prestigious schools.

Universities also played an important role in the establishment of examinations in Britain. London created a matriculation examination in 1838, which in 1842 became the earliest formal written school examination.⁹ The system established at the Society

of Arts, taken as an exemplar by other systems, was modeled on the written and oral examinations used at the University of Dublin. Oxford and Cambridge established systems of ‘locals,’ examinations graded by university “boards” to assess local school quality. In 1858, they began to use these examinations for individual students and, in 1877, to select them for university entrance. Other universities (Dublin and Durham) followed the same path and established procedures for examining local school pupils. The system of university control of examinations continued throughout the second half of the 19th century.

During the 18th and 19th centuries European countries also began to develop examinations for selection into the professional civil service. The purposes of the examinations were to raise the competency levels of public functionaries, lower the costs of recruitment and turnover, and control patronage and nepotism. Prussia began using examinations for filling all government administrative posts starting as early as 1748, and competition for university entrance as a means to prepare for these examinations followed. The British introduced competitive examinations for all civil service appointments in 1872.

Public examination systems in Europe, therefore, developed primarily for selection, and when mass secondary schooling expanded following World War II, entrance examinations became the principal selection tool setting students on their educational trajectories. In general, testing in Europe controlled the flow of young people into the varying kinds of schools that followed compulsory primary schooling. Students who did well moved on to the academic track, where study of classical subjects led to a university education; others were channeled into vocational or trade schools.

In the last two decades, the duration of compulsory schooling has become longer; the trend has

⁷The 12 members of the European Community (EC) are Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, The Netherlands, Portugal, Spain, and the United Kingdom. Much of the general discussion of EC education and examination systems is taken from Madaus and Kellaghan, op. cit., footnote 1. For comparative data on U.S. and Japanese education, see, e.g., Edward R. Beauchamp, “Reform Traditions in the United States and Japan,” William K. Cummings, Edward Beauchamp, Shogo Ichikawa, Victor N. Kobayashi, and Morikazu Ushioji (eds.) (New York, NY: Praeger Publishers, 1986).

⁸In the United States, secondary schooling is more closely linked, in structure and content, with primary than with university education. Other countries’ elite secondary schools are closely linked to universities. See Martin Trow, “The State of Higher Education in the United States,” in Cummings et al., op. cit., footnote 7, p. 177.

⁹Some professional bodies had already introduced written qualifying examinations (Society of Apothecaries in 1815 and Solicitors in 1835). The London examination initiated in 1842 was the first formal school examination of its kind.

generally been to provide access to comprehensive schooling for more students and to provide a wider variety of academic and vocational choices. Examinations that filter students into different kinds of schools, once given at the end of primary school (around age 11), now take place around age 16 or even 18. The uses and formats of these "school-leaving" examinations are evolving as more options have become available and larger percentages of students seek and can gain access to postsecondary education. In several countries, school-leaving examinations that were once considered a passport to higher education have evolved into first stage or qualifying examinations, which are followed by more diversified examinations for specific prestigious universities or lines of study administered by the university itself. Examples are the French *Baccalaureate*, the German *Abitur*, and the Japanese *Joint First Stage Achievement Test (JFSAT)*.¹⁰

Standardized examinations are not generally used outside the United States for purposes other than certification or selection. However, some exceptions are noteworthy. In Sweden, standardized examinations are used as scoring benchmarks to help teachers grade students uniformly and properly in their regular classes. Examination results in a few countries serve not only to evaluate student performance but also to evaluate the quality of a teacher or school. This was the approach, now abandoned, in England during the second half of the 19th century, when "payment by results" was based on student scores.¹¹ Today student scores in **China have taken** on this school accountability function, in that "Key Schools" in China receive extra resources in recognition of their better examination results.¹²

Central Curricula

In most EC countries curriculum is prescribed by a central authority (usually the Ministry of Education). However, the level of prescription varies from system to system. In Germany, curricula are determined by each of the 11 States,¹³ in France the curriculum is quite uniform nationwide, and in Denmark individual schools enjoy considerable discretion in the definition of curricula. The trend in several countries has been to allow schools a greater say in the definition of curricula during the compulsory period of schooling; school-based management and local control are not uniquely American concepts.

The United Kingdom¹⁴ seems to be moving in the other direction. In the past, curricula in the United Kingdom were determined by the local education authorities and even individual schools. Independent regional examination boards exerted a strong influence on the curricula of secondary schools. The central government significantly tightened its grip around the regional boards beginning in the mid-1980s, and since the Education Reform Act of 1988 the U.K. has moved toward adoption of a common national curriculum.

Divisions Between School Levels

Most European countries have maintained the conventional division between primary, secondary, and third-level education. The primary sector offers free, compulsory, and common education to all students; the secondary level is usually divided into lower and upper levels. The duration of primary schooling can vary among the States or provinces of a given country.

¹⁰This has changed slightly with the change from the Joint First Stage Achievement Test (JFSAT) to the Test of the National Center for University Entrance Examinations (TNCUEE). The JFSAT was required only for those candidates applying to national and local public universities (approximately 49 percent of total 4-year university applicants), not those applying to private universities. Some applicants for private universities now also take the TNCUEE. Shin' ichiro Horie, Press and Information Section, Embassy of Japan, personal communication, Aug. 2, 1991.

¹¹In 1862, the British Government adopted the Revised Code of 1862, which established the criteria for the award of government grants to elementary schools. Each child of 6 and over was to be examined individually by one of Her Majesty's Inspectors toward the end of each school year. Attendance records were also taken into consideration. Thus, each child over 6 could earn the school 4 shillings for regular attendance and a further 8 shillings for successful performance in the annual examination. Clare Burstall, "The British Experience With National Educational Goals and Assessment," paper presented at the Educational Testing Service Invitational Conference, New York, NY, October 1990.

¹²Eckstein and No@ op. cit., footnote 6, p. 307.

¹³This is also the case in Canada and Australia, where each of the provinces or States sets its own curricula.

¹⁴The term "United Kingdom" (England, Wales, Scotland, and Northern Ireland) is used throughout this document. Testing practice in Northern Ireland, England, and Wales is similar, but Scotland is unique, with a completely different structure of testing and examinations. Scotland has only one examining board, with close connections to the central Scottish Education Department; the other countries in the United Kingdom each have several examining boards. Desmond Nuttall, director of the Centre for Educational Research, London School of Economics and Political Science, personal communication June 1991.

Table 5-1-Data on Compulsory School Attendance and Structure of the Educational Systems in the European Community

	Comprehensive attendance (age)	Horizontal structure of system (years)	Compulsory curriculum/schools (lower secondary grades)	Differentiated Curriculum/schools (grades)
Belgium ^{a,b}	6-16 (16-18 P-T)	6-3-3 or 6-2-2-2	7-1 o'	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 ^b	6-15	4-6-3	5-6 ^f	5-13
Greece	6-15	6-3-3	7-9	10-12
Ireland ^a	6-15	6-3-2 or 3	7-9 ^c	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 ^e	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

^aBelgium and Ireland have an additional 2 years preprimary education integrated into the primary school system. All other countries have provision outside the formal educational system for early childhood education.

^bBelgium and Germany are federations. There are two States in Belgium with completely independent educational systems. There are 11 States in the former Federal Republic of Germany (16 in the new Germany). Each of the 11 States determines its curriculum under terms agreed by the Council of State Ministers of Education.

^cA number of countries are less advanced than others in comprehensiveness of their school structures.

SOURCE: George F. Madaus, Boston College, and Thomas Kellaghan, St. Patricks College, Dublin, "Examination Systems in the EuroDean Community: Implications for a National Examination System in the United States," OTA contractor report, April 1991, table 3.

Most European countries at one time required a national school examination at the end of primary schooling. These examinations were intended to clarify for teachers the standards that were expected, provide a stimulus to pupils, and certify completion of a phase of formal education. They were used for admission to secondary education and for pre-employment screening. But these examinations raised many concerns about their limiting effects on the curriculum and about the tendency among some schools to retain students in grade in order to prevent the low achievers from presenting themselves for examinations.

Perhaps most important, however, were the changes in the philosophy of education that led to raising the school-leaving age and provision of adequate space in secondary schools to accommodate all students. Secondary education was once highly selective, with relatively low participation rates beyond the primary level, and with major divisions between two or three types of schooling. The most exclusive was the "grammar school," "gymnasium," or "lycee," which prepared students for third-level education

and professional occupations. Typically, the school systems of Europe offered a classical academic curriculum in the liberal arts. As numbers of students in this line of study grew, the traditional academic curriculum became diversified, subjects were presented at different levels, and some students took practical or commercial-type subjects.¹⁵

After the second World War, and particularly during the 1960s, demographic, social, ideological, and economic pressures led to various reviews of education. All the EC countries have made some moves to provide comprehensive lower secondary education (up to age 15 or 16), but these patterns are varied (see table 5-1). Several countries have established comprehensive lower secondary school curricula. Denmark and Britain have gone the furthest, with 10 years of comprehensive education. Greece, Portugal, Spain, Italy, and France also have relatively long periods of comprehensive education. 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 tradi-

¹⁵The alternative to the academic secondary school were schools offering technical curricula to prepare students for skilled manual occupations. These schools also expanded their range of offerings as the numbers of students grew, but they typically provided practical, usually short-term, continuing education.

Table 5-2-Upper Secondary Students in General Education and in Technical/Vocational Education, by Gender, 1985-86 (In percent)

	Girls		Boys	
	General education	Technical/vocational education	General education	Technical/vocational education
Belgium ^a	56%	44%	53%	47%
Denmark.....	40	60	26	74
France ^b	65 ^c	35	58 ^c	42
Germany ^b	51	49	57	43
Greece.....	83	17	62	38
Ireland.....	79	21	86	14
Italy ^d	26	74 ^e	22	78 ^e
Luxembourg.....	38	62	29	71
Netherlands.....	49	51	43	57
Portugal ^f	99	1	99.8	0.2
Spain.....	58	42	53	47
United Kingdom.....	53	47	57	43

^aLower and upper secondary education.

^b1986-87.

^cIncludes upper secondary technological education.

^d1984-85.

^eIncludes preschool and primary teacher training.

^fTechnical/vocational education was abolished in 1976. New courses were introduced on an experimental basis in 19s3/64.

SOURCE: European Communities Commission, *Girls and Boys in Secondary and Higher Educational* (Brussels, Belgium: 1990), table 3b.

tional German school structure (the classical *gymnasium* and the vocational school) have been sufficiently strong and successful to resist possible merging. 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.¹⁶

Today the term “general education” is used to describe the activities of schools that include university-preparation curricula as well as programs designed for students who are not likely to go on to university. Nevertheless, the upper secondary level in all European countries is still quite differentiated, especially in Germany and Italy. (In Italy the system is so complicated that it has been described as a “jungle.”¹⁷) As shown in table 5-2, in 8 of the 12 EC countries a majority of students follow a curriculum of general education, but a sizable number of students are in technical/vocational education courses. Comprehensive high schools in the United King-

dom, France, and, to a somewhat lesser extent, Germany, have begun to resemble the typical comprehensive American high school.

These shifts toward comprehensive schooling have resulted in changed testing policies. Today none of the EC countries administers a national examination at the end of primary schooling.¹⁸

Variation in the Rigor and Content of Examinations

Specified examinations for leaving secondary school and moving into higher levels of schooling vary across locales, kinds of degrees, subject areas, and competitiveness of the program or of the university. For example, while the French *Bac* retains a large core of general education subjects that all candidates are required to take (albeit with different weights), the 4 options offered in 1950 had grown to 53 in 1988.¹⁹

¹⁶Madaus and Kellaghan, *op. cit.*, footnote 1, pp. 53-54.

¹⁷*Ibid.*, p. 55.

¹⁸*Ibid.* Note, however, that Italy uses school-based **primary** examinations set, administered, and scored by the pupils' own teachers. The United Kingdom has plans to introduce nationwide assessment at ages 7 and 11, but these will be scored by teachers and used for accountability, and are not intended to be used for selection. Some schools in Belgium also administer an **examination** at the end of primary schooling, but this is a local school option, not a national policy.

¹⁹Information about the *Bac* was provided to OTA by Sylvie Auvillain of the French Embassy, July 1991. See also the final section in this chapter for a more detailed discussion of the French examination system.

On the basis of examination performance, a candidate is usually awarded a certificate or diploma that contains information on performance on each subject in the examination in letters (A, B, C, D, E) or numbers (1, 2, 3, 4, 5). Usually, grades are computed by summing marks on sections of questions and on clusters of questions or papers. The final allocation of grades may also take into account grade distributions in previous years. These marks or grades are used in making university admissions decisions.

The certificate or diploma may also confer the right to be considered for (if not actually admitted to) some stratum of the social, professional, or educational world. Certificates are credentials, and certification therefore plays a dual role: educationally, in establishing standards of academic achievement, and socially, in justifying the classification of individuals into categories that determine their shares of educational resources and employment opportunities.

Because government manages and finances higher education, and scholarships often cover almost all university costs in some countries, stiff entry competition is seen as a fair and appropriate way to distribute scarce educational resources.

psychometric Issues

Two major criteria for European examinations are objectivity and comparability. The central concern is whether the examinations reflect what is in the syllabus and whether they are scored fairly. Since, as noted below, most of the examination questions are essay questions that cannot be machine scored, it is not surprising that these issues of fairness are foremost. In the United States, test fairness issues have been analyzed primarily through statistical methods. This statistical apparatus, known as psychometrics, has been honed over seven decades of research and practice. It attempts to identify item or test bias,²⁰ and determine the reliability and validity of tests. Although European educators attempt to ensure that examinations reflect what is in the syllabus (i.e., content validity) and whether they are scored fairly (i.e., reliability), they do not typically conduct intensive pretesting and item analysis;

quantitative models of item-response theory, equating, reliability, and validity receive little or no attention. Unlike the United States, Europe does not have an elite psychometric community with strong disciplinary roots, or an extensive commercial test industry.²¹ Only the United Kingdom has *made any* attempt to apply to their examinations psychometric principles of the type developed in the context of U.S. testing, and they are still not in widespread use.

Essay Format and the Cost Question

Because examinations in European countries require students to construct rather than select answers, the examinations are considerably more expensive to score than the multiple-choice tests common in the United States. (Multiple-choice tests, on the other hand, are relatively expensive to design. See ch. 6 for discussion.) In general, the more open-ended a test is, the more expensive it will be to score, since scoring requires labor-intensive human judgment as opposed to machine scoring. The achievement tests used in other countries typically assess mastery and understanding of a subject by asking students to write. A few require oral presentations (Germany, France, and foreign language examinations in many countries). Some of the German *Abitur* requires students to give practical demonstrations in subjects such as music and the natural sciences.

These tests are expensive-to grade them takes the time of trained professionals (teachers, examiners, university faculty, or some combination). For example, written examinations taken at age 16+ in Great Britain and Ireland cost roughly \$110 per student.²² (In Ireland, candidates pay about 40 percent of the cost.) These costs maybe tolerable in countries where a small percentage of the age cohort takes the examination. But in the United States, with nearly five times as many students in this age group, testing the 3 million 16-year-olds in U.S. schools using the British or Irish model would cost about \$330 million. Looked at from the perspective of one State, Massachusetts, it would cost almost \$7 million to test all 65,000 16-year-old-students using the model of essay on demand; at present, Massachusetts spends just \$1.2 million to test reading,

a recent Summary and discussion of the meanings of test bias see, e.g., Walter Haney, Boston College, "Testing and Minorities," draft January 1991. See 6 for an explanation of reliability, validity, and other psychometric concepts.

op. cit., footnote pp. 57-58.

²⁰Ibid., pp.

writing, and arithmetic achievements of students in three grades and three subjects.²³

An additional factor to be included in a cost analysis is the potential effect of tests on retention. In the United Kingdom, for example, many students remain in school an extra year to repeat the General Certificate of Secondary Education (GCSE) if they did not pass the first time, or to repeat the more advanced ‘‘A levels’’ if they wish to try for a higher grade.

Tradition of Openness

Individual test takers in the United States can request prior year examinations and sample examination booklets for some tests used for selection, i.e., the Scholastic Aptitude Test (SAT); in addition, third-party vendors offer test preparation classes or software to enable students to practice for these examinations. In general, however, there is a greater emphasis on test security in the United States than in other countries,²⁴ where both the examinations and correct responses are made public following an examination and become the subject of much discussion. In France, for example, examination questions make front page news, and in Germany, answer scripts are returned to students who may question the way they were graded with their teachers. If a problem cannot be resolved between the student and teacher, the matter is referred to the Ministry of Education.

In the United States, legal challenges since 1980 have made the disclosure of college admissions tests available to test takers who wish to review them, but the examinations are not routinely publicized as in Europe. Some observers contend that releasing examination questions helps focus student and teacher awareness on the facts, concepts, or skills required in order to do well on the test, and that ‘‘teaching to the test’’ is therefore a good thing. Multiple-choice examinations, however, which are quite inexpensive to score, are very costly to

develop, because of the time and effort spent pretesting items and attempting to eliminate various biases. Releasing such tests in advance, therefore, could jeopardize their validity; this is important because of the high costs of creating new items.

The Changing State of Examinations in Most Industrialized Countries

There have been important changes in European test policies in the past three decades; many of the most dramatic changes have been undertaken in the last few years. France abolished centralized examinations at age 16+ with the aims of postponing selection, making assessment more comprehensive, and giving a greater role to teachers in assessing students. However, the examinations were reinstated in the 1980s, at least partly because the resources to support a school-based system of assessment had not been made available to the schools.²⁵ The United Kingdom is overhauling its examination system. Even in Japan, where success in examinations has been the central feature of the educational experience, politicians and educators are debating and reevaluating the form and functions of national examinations.

A major force affecting examination policies has been expansion of the educational franchise. Rising participation rates and rising expectations of individuals with diverse ethnic and socioeconomic backgrounds have changed attitudes toward the assessment of student progress and the uses of tests for important economic and social decisions. Historical criticisms of the narrowing effects of these examinations on students’ educational experiences have become politically significant. Many commentators always judged tests unsuitable for low-achieving students, an argument that has gained credence in the light of data suggesting that in order to avoid the examinations these students are likely to leave school early and enter the labor force without

²³It should be noted that the United States has some experience with nationally standardized written examinations. The Advance Placement (AP) program for instance, includes tests comprised of short answer and essay items. Currently the AP test costs \$65 per subject per student, paid for in most cases by the student rather than the school system. This financial burden prevents some poor students from taking the tests required for college credit. Some States (Florida and South Carolina), pay all AP fees and others (Indiana and Utah) subsidize or help students in need, but most States have no official policy, although the Educational Testing Service reduces the fee to \$52 for those with need. Jay Mathews, ‘‘Low Income Pupils Find Exam Fees a Real Test: California Questions Who Should Foot the Bill,’’ *The Washington Post*, Apr. 25, 1991, p. A3.

²⁴Public Law 100-297, which authorizes the U.S. Secretary of Education to approve comprehensive tests of academic excellence, specifies that, besides being conducted in a secure manner, ‘‘. . . the test items remain confidential so that such items maybe used in future tests.’’ This law has been passed, but funding has not been appropriated.

²⁵Madaus and Kellaghan, op. cit., footnote 1, p. 60.

benefit of any formal certification.²⁶ The apparent correlation between participation rates and school-leaving examination policies is striking: in the United Kingdom, for example, the participation rate drops from almost 100 percent at age 15 to just under 70 percent at age 16—when examinations must be taken. In contrast, some 95 percent of all American 16-year-olds are still in school (see table 5-3).

As noted above, a second area where examination policies have changed is the elimination of standardized examinations at the primary level. Furthermore, at the secondary level there has been a move toward greater reliance on assessments developed and scored by teachers. In four EC countries (Belgium, Greece; Portugal, and Spain), national examinations have been abolished and certification is entirely school based at both primary and secondary levels. In other countries, teachers may mark examinations set by an outside body or contribute their own assessments, which are combined with the results of the standardized examinations. This was the pattern in Britain from the 1960s onward, and virtually every GCSE examination includes an assessment (of things like oral work, projects, and portfolios) by teachers. Although the national program is bringing more centralized curriculum to the United Kingdom, the national curriculum assessment relies extremely heavily on teacher assessments.²⁷

A third trend has been the shift in emphasis from selection to certification and guidance about future academic study. This shift has been made possible, especially at lower educational levels, by the expansion of places in secondary schools. Furthermore, as the examinations have become more varied, selection for traditional third-level education is no longer a concern for as many students. Increasing numbers are now turning to apprenticeships or technical training.

Other Considerations

There are other important variables that affect the administration, costs, and outcomes of testing. These include the numbers of students to be tested, preselection of students prior to testing, the homogeneity of the student population and of the teaching

Table 5-3—Enrollment Rates for Ages 15 to 18 in the European Community, Canada, Japan, and the United States: 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 ^a	100.0	94.8	81.7	67.8
(of whom, part-time) . . .			(0.1)	
Greece ^b	82.1	76.2	55.2	43.6
Ireland ^b	95.5	83.9	66.4	39.6
Italy	—	—	—	—
Luxembourg ^c	—	—	83.4	71.1
(of whom, part time) . . .			(15.8)	(15.8)
Netherlands ^d	98.5	93.4	79.2	59.7
Portugal		32.1	36.9	29.2
Spain	84.2	64.7	55.9	30.4
United Kingdom	99.7	69.3	52.1	33.1
Canada	98.3	92.4	75.7	56.9
Japan ^c	96.6	91.7	89.3	3.2
(of whom, part-time) . . .	(2.6)	(1.9)	(1.7)	(1.4)
United States ^b	98.2	94.6	89.0	60.4

^aApprenticeship is classified as full-time education.

^b1986-87.

^cExcluding third level.

^dExcludes second level part-time education.

SOURCE: George F. Madaus, Boston College, and Thomas Kellaghan, St. Patrick's College, Dublin, "Student Examination Systems in the European Community: Lessons for the United States," OTA contractor report, June 1991, table 5; information for this table from Organisation for Economic Cooperation and Development, *Education in OECD Countries, 1987-88* (Paris, France: 1990), table 4.2, except figures for Portugal which are for secondary education in 1983-84 and come from European Communities Commission, *Girls and Boys in Secondary and Higher Education* (Brussels, Belgium: 1990), table 1c.

profession, centralization and consistency of teacher training to support common standards, and the number of days in the school year. These issues need to be included in efforts to compare testing policies across countries. There is no one model that could be described as the European examination system and, more importantly, no one model that can be transplanted from its European or Asian setting and be expected to thrive on American soil.

Lessons for the United States

What lessons from European and Asian testing policies apply to the American scene? To address that question OTA focused attention on three basic

²⁶In Britain and Ireland, the number of such students are about 11 and 8 percent, respectively. *Ibid.*, p. 15. (This estimate appears low to other researchers. Max Eckstein, professor of Education, Queens College, City University of New York, personal communication, 1991).

²⁷Nuttall, *op. cit.*, footnote 14.

issues: the functions, format, and governance of testing.²⁸

Functions of Testing

This report concentrates on three basic functions of educational testing: instructional feedback to teachers and students, system monitoring, and selection, placement, and certification (see ch. 1). European and Asian testing systems, though different from country to country, tend to emphasize the last group of functions, i.e., selection, placement, and certification.²⁹ There is in other countries almost no reliance on student tests for accountability or system monitoring, activities that are typically handled through various types of ministerial or provincial inspectorates; this fact itself suggests an important lesson for U.S. educators.

Selection, Placement, and Credentialing

If one wished to import testing practices from overseas, an obvious strategy would be to expand and intensify the use of student testing for selection, placement, and certification decisions. Indeed, this appears to be at least one of the ideas behind some proposals for national achievement testing in the United States.³⁰ OTA finds that the European and Asian experience with testing for these functions leads to three important lessons for U.S. policymakers.

First, in most other industrialized countries, the significance of testing is greatest at the transition from secondary to postsecondary schooling. Standardized examinations before age 16 have all but disappeared from the EC countries. Primary certificates used to select students for secondary schools have been dropped as comprehensive education past the primary level has become available to all students. Current proposals for testing all fourth graders with a common externally administered and graded examination would make the United States

the only industrialized country to adopt this practice.³¹

Second, the continued reliance on student testing as a basis for allocating scarce publicly funded postsecondary opportunities has, in Europe and Asia, come under intense criticism. Having relatively recently attempted to relax stringent elementary and secondary school tracking systems, many countries have been reluctant to hold on to stiff examination-based criteria for admission to third-level schooling. As a result, admissions policies have been in flux. It would be ironic if U.S. policymakers, in an attempt to import the best features of other countries' models, adopted a system of increased selectivity—even at the postsecondary level—just when those countries were evolving in the other direction.

In this context it is important to note the fundamental differences in the relationships between secondary and postsecondary schooling in the United States and elsewhere. In most other industrialized countries, there is a strong link between secondary schools and the universities for which they prepare students; in the United States, on the other hand, high school graduates face a vast array of postsecondary opportunities, diverse in their location, academic orientation, and selectivity. Although periodically in American educational history there have been attempts to influence secondary school curricula and academic rigor through changes in college admissions policies, the postsecondary sector in the United States has remained basically independent of the system of primary and secondary public schools. Restructuring the linkages between these sectors along the lines of the European model, and changing the examination system accordingly, could bring about changes in the quality of American high school education; but the benefits of such a policy need to be weighed against the uncertain effects it would have on the U.S. postsecondary

²⁸This framework was suggested by Max Eckstein, professor of Education, Queens College, City University of New York, who chaired an OTA workshop on lessons from testing in other countries, January 1991.

²⁹Classroom testing, conducted by teachers to assess on a regular basis the progress of their students, is likely to be much the same around the world—teacher-developed quizzes, end-of-year examinations, and graded assignments do not vary much from Stockholm to Sacramento, from Brussels to Buffalo.

³⁰See, e.g., Madaus and Kellaghan, *op. cit.*, footnote 1, for an overview of national testing proposals. It should be noted that many advocates of high-stakes selection and certification tests view their principal role as stimulus to improved learning and teaching. Although this might be considered a fourth function of testing, this report treats the potential motivating effects of tests as a crosscutting issue affecting the utility of tests designed to serve any of the three main functions.

³¹As discussed earlier, the United Kingdom has implemented a new system of national assessment at ages 7 and 11, for purposes of accountability (system monitoring).

sector, considered by many to be the best in the world.³²

The third lesson concerns the equity effects of increased testing for what are commonly called “gatekeeping” functions. Europe has a long history of controlled mobility among nations, and an equally long history of efforts to deal with changing ethnic and national composition of its population. What is relatively new in many countries, however, is the commitment to widening educational and economic opportunities for all citizens. As a result of this shift in social and economic expectations, the use of rigorous academic tests as gatekeepers has come under fire in many countries. In France, for example, the expansion of options under the *Bac* emerged from the struggle of the 1960s to reform not only the schools but much else in French society.

In discussions with many educators and policymakers from European countries, OTA found a fairly common and growing concern with the equity implications of educational testing; European (and to a lesser extent Asian) education policymakers are in fact looking to the United States for lessons about how to design and administer tests fairly. Although the ultimate resolution of complex equity issues escapes predictability, there is no doubt that continued cross-cultural and translational exchanges among policymakers and educators grappling with these issues will be invaluable.

System Monitoring

European and Asian nations tend not to use student examinations to gauge the performance of their school systems. That function is still handled primarily by inspections carried out at the ministerial or provincial government levels. There has been heightened interest in using the results of international comparative test score data for policymaking, although exactly how to use the data for internal policy analysis is a relatively new question.³³ Nevertheless, three lessons for the United States emerge from the European and Asian experiences.

First, other countries considering the adoption of some kind of test-based accountability system tend to view the American National Assessment of Educational Progress (NAEP) as a model. The fact that NAEP uses a sampling methodology, addresses a relatively wide range of skills, and is a relatively “low-stakes” test make it appealing as a potential complement to other data on schools and school systems. One lesson for American policymakers, therefore, is to approach changes to NAEP cautiously (see also ch. 1 for a thorough discussion of NAEP policy options).

The second lesson is to consider nontest indicators of educational progress that could be valid for monitoring the quality of schools. In this regard, careful study of the ways in which inspectors operate in other countries—how they collect data, what kind of data they collect, how their information is transmitted, how they maintain neutrality and credibility—could be fruitful.³⁴

Finally, the European and Asian approach to system monitoring suggests a general caution regardless of whether tests, inspections, or other data are utilized. Public perception of the adequacy of schools in most countries depends on which schools are in question: parents typically like what their own children are doing, but complain about the system as a whole. It is difficult to pinpoint the causes of this dual set of attitudes;³⁵ in any event, it is fairly clear that there is greater enthusiasm for reform in general than for changes that might affect one’s own children. Like the ‘not-in-my-back yard’ (‘NIMBY’ problem faced by environmental policymakers, education policymakers in many countries face a formidable “NIMSY” problem: education reform may be OK, so long as it is “not-in-my-school yard.” American, European, and Asian educators and policymakers who have struggled with the NIMSY problem in their attempt to respond effectively to analyses of various types of system monitoring data could learn much from one another.

³²See Kirst, *op. cit.*, footnote 4, for discussion of the quality of U.S. colleges and universities.

³³The Organisation for Economic Cooperation and Development (OECD) has been sponsoring, along with the U.S. Department of Education, an ongoing collaborative effort to better understand and utilize comparative data on student achievement.

³⁴For discussion of multiple indicators of education, see U.S. Department of Education, National Center for Education Statistics, *Education Counts: An Indicator System to Monitor the Nation’s Educational Health* (Washington DC: 1991).

³⁵One explanation that caused a stir in policy circles was the finding that statewide achievement scores in every State were above the national average. See discussion in ch. 2 of this report.

Test Format

In European countries, the dominant form of examination is “essay on demand.” These are examinations that require students to write essays of varying lengths. Use of multiple-choice examinations is limited, except in Japan, where multiple-choice tests are common at all levels of elementary and secondary schooling and are used as extensively as in the United States. Performance assessments of other kinds (demonstrations, portfolios) may be used for internal classroom assessment, but not generally for systemwide examinations because of costs.

The lesson from this mixture of test formats overseas is a complicated one. On the one hand, European experience could lead American policymakers to eliminate, or at least reduce significantly, multiple-choice testing; surely some critics of U.S. testing policy would embrace this position. But this inference would be erroneous, given the conflicting evidence from the overseas examples. For example, if one of the purposes of testing is to raise standards of academic rigor, the French and Japanese examples offer conflicting models: both countries typically rank higher than the United States in comparisons of high school students’ achievement, but they rely on diametrically different methods of testing.

If there is a lesson, then, it is that testing in and of itself cannot be the principal catalyst for educational reform, and that changes in test format do not automatically lead to better assessments of student achievement, to more appropriate uses of tests, or to improvements in academic performance. The fact that European countries do almost no multiple-choice testing is not, in itself, a reason for the United States to stop doing it; rather it is a reason to consider whether: a) reliance on the multiple-choice format satisfies the numerous objectives of testing; and b) whether alternative formats in use in other countries, such as essays and oral examinations, could better serve some or all objectives of testing in the United States.

In considering alternative test formats and the experience of other countries, it is important to keep two additional issues in mind. First, as discussed in chapters 4 and 8 of this report, the combination of multiple-choice and electromechanical scoring tech-

nologies made the concept of mass testing in the United States economically feasible. To the extent that this type of testing went hand in hand with the American commitment to schooling for all, it will be interesting to observe whether increased efficiency of test format will evolve as an important consideration in European countries committed to expansion of school opportunities for the masses.

Second, one of the important advantages of the multiple-choice format is that tests based on many different questions are usually more reliable and generalizable than tests based on only a few questions or tasks.³⁶ It allows for statistical analysis of test reliability and validity both before and after tests are administered. In addition, multiple-choice tests allow for statistical analysis of items and student responses, not as easily accomplished with performance assessments. If criteria such as reliability and validity remain a central concern among American educators, the adoption of European testing methods will necessitate substantial investments in research and development to bring those methods up to acceptable reliability and validity standards.

Governance of Testing

None of the countries studied by OTA has a single, centrally prescribed examination that is used for all three functions of testing. Moreover, the countries of Europe and Asia exhibit considerable variation in the degree of centralized control over curriculum and testing. In some countries, there are centrally prescribed curricula that are used as a basis for the standardized examinations students take, while elsewhere decisionmaking is more decentralized. An obvious lesson, then, is that the concept of a single national test is no less alien in other countries than it has been in the United States. Nevertheless, there are important differences in the governance of tests between the United States and other industrialized countries.

Testing and Curriculum

Although most countries allow some local control of schooling, in general there is greater national agreement over detailed aspects of curriculum than there is in the United States. This sense of a shared mission is reflected in tests that probe content mastery at much deeper levels than most of the

³⁶See discussion of generalizability in ch. 6.

standardized tests in the United States.³⁷ As explained elsewhere in this report, however, this has more to do with the politics of testing than with the technology of testing: the United States has a long history of decentralized decisionmaking and school governance, and an aversion to the idea of curricula defined for the Nation as a whole. Standardized tests that can be used across the United States have therefore been limited to skills and knowledge common to most school districts—which has meant basic reading, writing, and arithmetic.³⁸ The pursuit of consensus in the United States for anything beyond the basics has proved difficult, though not impossible; the best example to date is NAEP, considered by most educators who are familiar with it as an important complement to the kinds of information provided on nationally normed standardized tests. Nevertheless, even NAEP items fall short of the complexity, depth, and specificity of content material attained in written examinations overseas.

Three important lessons regarding governance of tests emerge for U.S. policy. First, consensus on the goals and standards of schooling appears easier to establish in Europe and Asia than in the decentralized and diverse U.S. education system. As a consequence, national examinations in Europe and Asia can be very content and syllabus specific. In the United States, on the other hand, achieving national consensus usually means limiting examinations to basic skill areas common to 15,000 school districts. Even NAEP, which consists of items derived from elaborate consensus-seeking processes, does not assess achievement at a level of detail and complexity comparable to typical essay examinations in other countries. The lesson from abroad, then, is that syllabus-specific tests can be national only in countries where curriculum decisions are made centrally or where consensus can be easily attained.

The second lesson, related to the first, concerns the sequencing of curriculum and test design. European and Asian experience does not demonstrate that national testing raises the academic rigor

of curricula, but rather that national consensus on goals and standards of schooling allows for consistent curricula that can be tested by syllabus-based national examinations. Indeed, the importance of keeping the horse of curriculum and instruction before the cart of assessment (one of OTA's central findings in this report) is reinforced by the overseas experience.

The third lesson concerns the effects of heavily content-driven examinations on student behavior. Syllabi, topics, criteria of excellence, and questions from prior examinations are widely publicized in other countries, where preparing for tests is encouraged. This emphasis on curricular content conveys an important signal to students in Europe and Asia: "study hard and you can succeed." In the United States, students are encouraged to work hard, but their success in gaining admission to college or in finding good jobs often depends on many other factors besides their performance on tests closely tied to academic courses they have taken. While there is clearly a need for tests that can assess fairly the differences in knowledge and skills of individuals from vastly diverse and locally controlled school environments,³⁹ there may also be considerable merit in the use of examinations that reinforce the value of studying material deemed worthy of learning.⁴⁰

The Private Sector

Only in the United States is there a strong commercial test development and publishing market. The importance of this sector, in terms of research, development, and influence on the quality and quantity of testing, cannot be overstated. Even when States and districts create their own tests, they often contract with private companies. In Europe and Asia, testing policies reside in ministries of education.

There is a certain paradox about the preference for public administration of tests in other countries and private markets in this country. Given that European and Asian countries typically have less trouble than

³⁷See, e.g., National Endowment for the Humanities, *National Tests: What Other Countries Expect Their Students to Know* (Washington, DC:1991), for examples of test questions faced by students in Europe and Japan.

³⁸For discussion of how multiple-choice items can assess certain "lower order thinking skills" see ch. 6.

³⁹See Donald Stewart, "Thinking the Unthinkable: Standardized Testing and the Future of American Education," speech before the Columbus Metropolitan Club, Columbus, OH, Feb. 22, 1989.

⁴⁰This issue turns on distinctions between aptitude testing and achievement testing (see ch. 6). For discussion of the historical development of these approaches to testing, see ch. 4. See also James Fallows, *More Like Us* (Boston, MA: Houghton-Mifflin, 1989), pp. 152-173.

the United States in defining national goals and standards of education, the ability to specify testing needs and contract with private vendors for test development and production ought to be relatively easier in other countries than in the United States. On the other hand, given that fragmentation in curricular standards and educational goals in the United States raise formidable barriers to market transactions, one might expect greater reliance on nonprofit or governmental organization of testing.

The Role of Teachers

Considerable responsibility is vested in teachers in other countries for the administration and scoring of standardized examinations. This practice is based on the premise that examinations with heavy emphasis on academic content should be developed and graded by professionals charged with delivering that content and respected for their ability to ascertain whether children are learning it. The important lesson for U.S. testing policy, then, is that faith in the professional caliber of teachers is a necessary condition for a credible system of examinations that requires teachers' judgments in scoring.

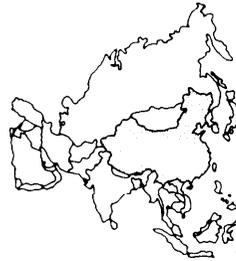
It is important to note that many European countries have only one or very few teacher training institutes, guaranteeing more consensus on the principles of pedagogy and assessment than in the United States, where teacher education occurs in thousands of colleges and universities. The centralized model of teacher training in other countries reinforces the professional quality of teaching, and makes it relatively easier to implement national curricula. The American tradition emphasizes standardized testing as a source of information to check teachers' judgments and to assure that children in diverse schools and regions are being treated equitably. The lesson from the European model, then, is that a centralized system of teacher preparation can increase the homogeneity of teaching and curricu-

lum and reduce the need for assessments designed to assure that all children are receiving similar educational experiences. This suggests a familiar theme: changing testing will not necessarily improve teaching, but changes in teaching can lead to different approaches to testing.

U.S. policymakers wishing to adopt examinations on European or Asian models will need to balance the need for increased reliance on teacher judgments with public demand for a system that provides an independent "second opinion," especially when test results have high stakes.

Snapshots of Testing in Selected Countries⁴¹

The People's Republic of China



The first examinations were attributed to the Sui emperors (589-618 A. D.) in China. With its flexible writing system and extensive body of recorded knowledge, China was in a position much earlier than the West to develop written examinations.

The examinations were built around candidates' ability to memorize, comprehend, and interpret classical texts.⁴² Aspirants prepared for the examinations on their own in private schools run by scholars or through private tutorials. Some took examinations as early as age 15, while others continued their studies into their thirties. After passing a regional examination, successful applicants traveled to the capital city to take a 3-day examination, with answers evaluated by a special examining board appointed by the Emperor. Each time the examination was offered, a fixed number of

⁴¹In the following country profiles all data on area and total population come from Mark S. Hoffman (ed.), *The World Almanac and Book of Facts, 1991* (New York, NY: Pharos Books, 1990); age of compulsory schooling and total school enrollment figures come from the United Nations Educational, Scientific and Cultural Organization (Unesco), *Statistical Yearbook* (Louvain, Belgium: 1985 and 1989). School enrollment figures include "pre-first level," "first level," and "second level" students. Data on number of school days comes from Kenneth Redd and Wayne Riddle, Congressional Research Service, "Comparative Education: Statistics on Education in the United States and Selected Foreign Nations," 88-764 EPW, Nov. 14, 1988.

For comparison purposes, current U.S. data are: size, 3.6 million square miles; population, 247.5 million. Mark S. Hoffman (ed.), *The World Almanac and Book of Facts, 1990* (New York, NY: Pharos Books, 1989). School enrollment: 46.0 million. U.S. Department of Education, National Center for Education Statistics, *The Condition of Education, 1991*, vol. 1, *Elementary and Secondary Education* (Washington DC: U.S. Government Printing Office, 1991).

⁴²Stephen P. Heyneman and Ingemar Fagerlind, "Introduction," in The World Bank, *University Examinations and Standardized Testing* (Washington, DC: 1988), p. 3.

Size	3,705,390 square miles, slightly larger than the United States
Population	1,130,065,000 (1990)
School enrollment	177.8 million (1988)
Age of compulsory schooling	6 to 16
Number of school days	September 1 to mid-July-- exact number of days not available
Selection points and major examinations	1. Provincial examinations at end of 9th year of compulsory schooling 2. Central examinations set by the State for university and college entrance
Curriculum control	National, central control

aspirants were accepted into the imperial bureaucracy.⁴³

Education in China today is largely centrally controlled. Curricula and the examinations that accompany them are used as a reflection of political philosophy and as a means of maintaining cultural cohesion, as well as to reinforce common loyalties in a population of over 1 billion people, speaking several major languages, distributed over a huge land mass (larger than the United States). There remains a sharp separation between academic schooling and vocational schooling, and examinations are the basis for making these selections at the end of the 9 years of compulsory schooling. Students may then enter general academic schools, vocational or technical schools, or ‘key schools,’ which accept the top cadre of students and receive superior resources in part based on the test results of their students. The examinations at this level are prepared by provincial education bureaus and are administered on a city-wide basis.

At the end of upper secondary school, students seeking university entrance take a centralized examination that provides no choice of subjects, specializations, or options. This examination is developed by the National State Education Commission and administered by provincial higher education bureaus who assign candidates to schools based on scores, specialties, and places available. The same is true for

technical schools. The Central Ministry of Labor and Personnel develops and *administers* a nationwide entrance examination for skilled worker schools. Strict quotas are assigned for overall opportunities for further study and to particular programs at specific institutions, based on a master plan of national and regional development goals. The size, wealth, and general power of certain municipalities (Beijing, Shanghai, and Tientsin) have enabled them to assume control over the examination mechanism, which in other locations may be directed by the central or provincial authority.

The number of candidates for university entrance is huge—in 1988, 2.7 million students prepared for the national college admission test. Less than one-quarter were accepted for study. Overall, about 2 percent of Chinese first graders eventually go on to higher education.⁴⁴ The format of the examinations, once extended answer/essay format, is beginning to change to short-answer and multiple-choice questions. Nevertheless, examinations are still scored by hand rather than machines. Some analysts suggest that, given the huge numbers of examinees, it is only a matter of time before machine-scorable formats are introduced, reinforcing the already strong emphasis in Chinese schools on rote learning and recall of facts.⁴⁵

The pendulum of Chinese higher education admission policy has swung with political pressures. After 1,000 years and a well-established tradition of using examinations to control admission to higher education and further training, the Chinese abolished examinations during the cultural revolution, with the goal of eliminating status distinctions. Selection was to be based instead on political activism and ‘correctness’ of social origin. The pendulum swung back again with the new regime in 1976, when examinations were reestablished as a means of allocating university places on basis of merit. Student scores rather than political orthodoxy have again become the major criterion to advancement. Examinations confer status in China. It is not uncommon to inquire about a person’s status in

⁴³William K. Cummings, ‘Evaluation and Examination,’ *International Comparative Education Practices: Issues and Prospects*, Thomas Murray (ed.) (Oxford, England: Pergamon Press, 1990), p. 90.

⁴⁴Harold J. Noah and Max A. Eckstein, ‘Tradeoffs in Examination Policies: An International Comparative Perspective,’ *Oxford Review of Education*, vol. 15, No. 1, 1989, p. 22.

⁴⁵*Ibid.*

society by asking: “How many examinations has he (or she) passed?”%

The Union of Soviet Socialist Republics (U. S. S. R.)⁴⁷



Soviet society has been characterized by central control and planning, and this centralization extends to the educational system.⁴⁸ The 15 republics and subrepublics that made up the U.S.S.R. had shared a central curriculum and common

school organization. Considerable local discretion had been provided, however, in education policy as it pertained to the secondary school-leaving certificate, the *attestat zrelosti* (maturity certificate). This certificate was based on accumulated course grades and an examination that was predominantly oral in nature. Each of the 15 republics was responsible for setting the content and standards of the examination, and the teachers who prepared the students dominated the process of setting the questions and evaluating the responses.⁴⁹

Because there was so little comparability in grading, the value of the *attestat zrelosti* meant different things in different parts of the country. As a result of this variability, the VUZy (universities and technical institutes) developed their own entrance examinations. Much like in the Japanese system, each university set its own questions, testing schedule and policy, cutoff score, and grading procedure. This diversified system placed a burden on students, who needed to negotiate a web of uncoordinated examinations, and travel great distances to sit for the necessary examinations at the university or institute of their choice. Much of the examination process involved oral examinations. The system was described as erratic, inconsistent, confusing, and subject to influence peddling and

Size	8,649,496 square miles, the largest country in the world, approximately 2.5 times the size of the United States
Population	290,939,000 (1990)
School enrollment	4.9 million (1988)
Age of compulsory schooling	7 to 17
Number of school days	September 1 to May30—exact number of days not available
Selection points and major examinations	1. Secondary school-leaving examinations set by each republic, graded by local teachers 2. Each university and technical institute sets its own entrance examination
Curriculum control	National, central control

corruption. There were persistent reports of discrimination against ethnic and religious groups in the examination process.⁵⁰

Controlling the flow of students into the university system was part of the overall regional and national planning that had been carried out through test quotas. During the revolution of 1917, university entrance examinations were abolished, and access was opened to all students. However, the examinations were reinstated in 1923.⁵¹ The more recent balance between central planning and local flexibility was another example of the need for political compromise. Some maintained that the tradeoff for local flexibility had been an incoherent and inconsistent system. In part to find more objective and standardized forms of testing, Soviets had begun looking to “American tests,” machine-scorable multiple-choice tests, for possible use in the *attestat zrelosti*. It is not clear how the various republics will react to relinquishing some of their local discretion in developing and scoring tests. As noted above, it is yet to be seen how the independence of the Soviet republics will affect the examination systems that were developed to serve the centralized political system of the past.

⁴⁶Eckstein and Noah, op. cit., footnote 6, p. 308.

⁴⁷This snapshot refers to the period before the recent breakup of the U.S.S.R. into separate republics.

⁴⁸Education and examination processes are undergoing radical changes and it is too soon to draw final conclusions. V. Nebyvaev, third secretary, Embassy of the Union of Soviet Socialist Republics, personal communication, July 31, 1991.

⁴⁹Noah and Eckstein, op. cit., footnote 44, p. 23.

⁵⁰Ibid.

S] Ibid.

Japan



When the United States compares itself to Japan, it is common to bemoan the fact that our schools are not more like theirs. Interestingly, one of the few things the two education systems have in common is their reliance

on machine-scorable multiple-choice examinations. In other ways our cultures and traditions are so different that many comparisons are superficial and, in some cases, potentially destructive.⁵²

When Japan emerged from its feudal period in the mid- 19th century, it began to look to the West for models to modernize aspects of Japanese life.⁵³ Among these models were the Western goals of compulsory primary education and of a high-quality university system. Japan also followed the French example of a centrally prescribed curriculum and textbooks, frequent testing during a school year, and end-of-year final tests. However, since Japanese students often finished the prescribed curriculum before the end of the school year, they began to focus on the use of entrance examinations for the higher level, rather than school-leaving examinations from the lower level. These entrance examinations became valued for several reasons. The first and most obvious was the need to select a few students from the many seeking higher levels of education. Another reason for devotion to examinations came from the uniquely Japanese cultural disposition known as *ie* psychology, “. . . the tendency to rigorously evaluate individuals before permitting them to join a family system or a corporate residential group, but once they are admitted, to accept and adjust to them as full members.”⁵⁴ This concept of first passing rigorous scrutiny and then receiving what becomes lifetime acceptance into established groups can be seen in acceptance of spouses into a family unit or employees into membership in Japanese firms.⁵⁵

Size	145,856 square miles, slightly smaller than California
Population	123,778,000 (1990)
School enrollment	21.2 million (1988)
Age of compulsory schooling	6 to 15
Number of school days	243
Selection points and major examinations	<ol style="list-style-type: none"> 1. Examinations for entry to some junior high and high schools 2. Joint First Stage Achievement Test: national preliminary qualifying examination for national local public universities (approximately 49 percent of all university candidates); abolished in 1989 and replaced with Test of the National Center for University Entrance Examinations for public universities (and some private universities) 3. Each university sets own College Entrance Examinations
Curriculum control	National, central control

The second major reform in Japanese schooling was implemented by the American occupation following World War II.⁵⁶ The School Education Law of 1947 caused a massive reorganization of the existing school facilities that is the basis for today’s educational system. Among these reforms were the establishment of a 6-year compulsory primary school and 3 additional years of a compulsory middle or lower secondary school. The first 9 years of compulsory education are free to all students. An additional 3 years of high school are modeled on the lines of the American comprehensive high school; however, all high schools charge tuition. While the law said that “. . . co-education shall be recognized in education,” many private junior high or high schools and some national and public local high schools are for one gender.⁵⁷

Higher education also was to be reformed, with the aim of broadening goals, leveling the traditional

⁵²See, e.g., Fallows, op. cit., footnote 40.

⁵³While the education system imported the “practical” disciplines (mathematics, science, and engineering) from the West, its moral content was strictly Japanese. The 1890 Imperial Rescript on Education made “the teachings of the ancestors of the Imperial Family” the basis for all instruction. “Education Reform in Japan: Will the Third Time be the Charm?” *Japan Economic Institute Report*, No. 45A, Nov. 30, 1990, p. 2.

⁵⁴William K. Cummings, “Japan,” in Murray (ed.), op. cit., footnote 43, p. 131.

⁵⁵Ibid.

⁵⁶“Education Reform in Japan,” op. cit., footnote 53.

⁵⁷Article 5 of the Fundamental Law of Education, Horie, op. cit., footnote 10.

hierarchy, expanding opportunities, and decentralizing control. While many of the reforms envisioned for changing higher education were not long-lived, opportunities were vastly expanded, and important powers devolved to universities, e.g., power over academic appointments, admissions, and so on. The postwar constitution formally guarantees academic freedom, and university autonomy is held sacred. Nevertheless, the government controls the purse strings for national universities, and ties between large employers and the national universities have led to a perpetuation of the hierarchy in Japanese education.⁵⁸

Japanese education today is highly centralized, with a common curriculum and little choice in subjects. Test scores become important early and throughout the structured progression of students along a carefully defined path. Some suggest this has had the impact of transforming Japan from an aristocracy to a society where what counts is the university one attends.⁵⁹ There is a progression, based on examinations, that has provoked considerable competition among students and their parents. While primary schools are quite egalitarian, many students compete for the more elite national junior high schools that grant entrance based on test scores and, in some cases, a lottery. There are also many private junior high schools whose entrance examinations are very competitive. It is hoped that success in an elite junior high will help guarantee entrance to the best high schools. There is space for approximately 60 percent of all the students in public high schools; private schools receive the rest.⁶⁰

Since there is now room for all students to attend high school of some sort, and since the curriculum is centralized, based on the university entrance examinations, today there is somewhat less competition for high school entry than in the past. But those high schools (public and private) with larger numbers of successful university applicants are still prized. Student selection to high school is based on prior grades and teacher recommendations as well as the high school entrance examination. With recent

education reforms, some of the pressure of this first stage of Japan's examination system has been reduced.

While the entrance examination system for Japanese universities has been in existence for over a century, the pendulum of common examinations v. university-developed examinations has swung back and forth. In the prewar period, an entrance examination was used only for those prestigious national universities that attracted large numbers of applicants. The private institutions did not require these examinations. With the postwar educational reforms, a single common examination, the Japanese National Scholastic Aptitude Test, was instituted for all universities. This examination was abolished in 1954 and replaced by a system whereby each university conducted its own entrance examination. School grades and recommendations from high school teachers were not given much weight, and eventually educators became concerned that the university entrance examinations did not adequately cover the scholastic ability of applicants.⁶¹

In 1979, therefore, a new system was put into place that eventually led to today's two-tiered examination system. The first stage required all applicants to national and local public universities (currently approximately 49 percent of all 4-year college applicants⁶²) to take the Joint First Stage Achievement Test (JFSAT), a retrospective examination created by the Ministry of Education. This examination was offered once a year to test mastery of the five major subjects in secondary school curriculum. In 1990, the JFSAT was abolished and replaced by the Test of National Center for University Entrance Examinations (TNCUEE). The main difference between these two tests is use and content. The JFSAT was required of applicants to national and local public universities only, while the TNCUEE is taken by some applicants for some private universities as well. In addition, the TNCUEE requires applicants to take examinations only in those subjects required by the universities to which

⁵⁸William Cummings, Harvard University, personal communication, August 1991.

⁵⁹In the United States and Korea, having the credential or degree is what counts in terms of prestige and career possibilities. In Japan, though, the status stems from attending a university: it is more important to be "Todai Man" —to attend Tokyo University, than to earn a Ph.D. James Fallows, personal communication, July 18, 1991.

⁶⁰Cummings, op. cit., footnote 58.

⁶¹Ikuo Amano, "Educational Crisis in Japan," in Cummings et al. (eds.), op. cit., footnote 7, pp. 38-39.

⁶²Horie, op. cit., footnote 10.

they are applying.⁶³ The second tier of examinations is the College Entrance Examinations (CEE), individually developed, administered, and graded by the faculties of each of the prestigious and highly selective universities.

While 34 percent of high school graduates seek university entrance, only 58 percent of these applicants gain entrance.⁶⁴ One-third⁶⁵ of the applicants each year are *ronin*, “masterless samurai,” who are repeating the examinations after attending special prep schools (*yobiko*) and *juku* (tutorial, enrichment, preparatory, and cram schools) in order to get higher scores, qualifying them for admission into the prestigious universities.

In fact, the *juku*, or cram school, and the *yobiko* have become almost a parallel school system to the public schools. The sole curriculum of these after-hours or additional schools is examination preparation. There are 36,000 *juku* in Japan. It is a \$5-billion a year industry. More than 16 percent of the primary school children and 45 percent of junior high students attend *juku*,⁶⁶ even though the extra schooling costs several hundred dollars a month and represents a significant financial burden for many families.⁶⁷ In fact, with competition even to gain entry into some of the most successful cram schools, some of which give their own admission tests, there are jokes about going to *juku* for *juku*.

There has been a great deal of concern in Japan about the impacts of “exam hell” in two regards—the impact on students and the impact on curriculum. In Japan, high school is not the time of exploration and discovery, socialization and extracurricular activities, football games and dating that is found in the American high school. Instead, students spend almost every waking hour in school, in *juku*, or at home studying. The school day is long and after school children go to *juku*; the school week extends through Saturday morning, and the school year is approximately 240 days long. Pressure is great and continuous until a student makes the final cut—

entrance into a prestigious university. One popular saying is: “Sleep four hours, pass; sleep five hours, fail.”⁶⁸

Other impacts are more subtle, but of equal concern: students who memorize answers but cannot create ideas, and a curriculum that focuses everything on preparation for the examinations. When students view schooling as “. . . truly relevant when it promotes preparation for the CEE and as only marginally useful when it does not contribute directly to university admission,”⁶⁹ this has a major cognitive and motivational impact on students’ approaches to education. It is not clear whether a love of learning for learning’s sake can be inspired later, once the student jumps the final hurdle and makes it to the home stretch of the university. Indeed, once accepted into college, students can take it easy and relax, discover the joys of the opposite sex and perhaps begin to rediscover some of the pleasures forsaken in their “lost childhoods.” In fact, the college period in Japan has often been referred to as a “4-year vacation,” although a well earned one, since the average Japanese student ranks at the top of the list in mathematics, science, and a number of other subjects in international comparisons.⁷⁰

France



The locus of control for education in France is the Ministry of Education (MOE). The curriculum, topics for examinations, and guidelines are set by MOE, with examination questions and overall administration coordinated by the 32 regionally dispersed academies. The Minister of Education sets a general program of what should be examined, but each academy is responsible for

⁶³Ibid.

⁶⁴Ibid.

⁶⁵Ibid.

⁶⁶Carol Simons, “*They Get by With a Lot of Help From Their Kyoiku Mamas,” *Smithsonian*, vol. 17, March 1987, p. 49.

⁶⁷Fallows, op. cit., footnote 59.

⁶⁸Simons, op. cit., footnote 66, p. 51.

⁶⁹Nobuo Shimahara, “The College Entrance Examination Policy Issues in Japan,” *Qualitative Studies in Education*, vol. 1, No. 1, 1988, p. 42.

⁷⁰Ibid., p. 52.

Size	220,668 square miles, about twice the size of Colorado
Population	56,184,000 (1990)
School enrollment	9.6 million (1988)
Age of compulsory schooling	6 to 16
Number of school days	185
Selection points and major examinations	<ol style="list-style-type: none"> 1. State-controlled brevet at end of comprehensive school (age 15) 2. <i>Baccalauréat</i> at completion of lycee (age 18), 38 options, 3 types of diploma, set by each regional academy with Ministry of Education (MOE) oversight 3. Admission to selective <i>grandes écoles</i> via <i>concours</i> after 1 to 2 more years
Curriculum control	National, central MOE control

administering the curricula and testing within a region.⁷¹

French students spend 5 years in the *ecole primaire*, or primary school, and move to the secondary school without taking a graduation or selection examination. However, there has been a recent interest in examining students to see how well the schools are doing. At the beginning of the 1989 school year, MOE, concerned with reports showing a large proportion of students (30 percent) with reading problems on entering secondary school, set out on an ambitious national examination that could be compared with the U.S. NAEP.⁷² Inspectors, teachers, and specialists from all across France gathered and created a matrix of national goals and achievement levels. Teachers submitted ideas for questions and, after a period of pretesting, the group developed a common standardized test for mathematics, reading, and writing at the third and sixth grade levels. All 1.7 million students in these grades were tested in their classrooms, and teachers administered and scored the tests using coded answer sheets. Since the goal was to diagnose individual problems, every student was tested and the results were sent to parents. Each teacher was given copies

of the exercises (a mixture of open-ended and multiple-choice questions) with discussion of the objectives, commentary on kinds of responses students made, and overall scoring results. Although summative national results were collected, there was to be no classification or comparison made between classrooms, schools, and regions. A followup to this examination was planned for September 1991, using a sampling of students rather than an every student census.⁷³

Democratic reform implemented some 15 years ago has meant that almost all 11-year-olds begin sixth grade in comprehensive secondary schools (college) of mixed ability levels. At the completion of comprehensive school, examinations for the *brevet de college* (college certificate) are given in three subjects: French, mathematics, and history/geography. The *brevet* examinations were abolished in 1977 and completely replaced by a school-based evaluation. However, because of concern with declining results and complaints about what it meant to complete secondary school, the *brevets* were reestablished in 1986. At present, graduation from secondary school is based on a combination of examinations controlled by the State and an evaluation by the school.⁷⁴

A common curriculum has been an expression of the value placed on the ideal of a unitary, cohesive, clearly defined French culture. Some have suggested this unity was won at the price of official neglect of minority and regional cultures within the country.⁷⁵ But this is changing, and nowhere is this change better reflected than in the discussion of what subjects should be taught at the *lycee* (the third level of schooling) and for the *Baccalauréat* (*Bac*), taken at the completion of the lycee. While once the focus was to provide the French *culture generale*, a common French culture through a central curriculum for the few who could demonstrate a high level of formal academic ability in literature, philosophy, and mathematics, this attitude has changed dramatically in recent years.

⁷¹Henk P.J. Kreeft (ed.), "Issues in Public Examinations," paper prepared for the International Association for Educational Assessment, 16th International Conference on Issues in Public Examinations, Maastricht, The Netherlands, June 18-22, 1990.

⁷²Marten Le Guen and Catherine Lacronique, "Evaluation CE-6ème. A Survey Report of Assessment Procedures in France on Mathematics, Reading and Writing," paper prepared for the International Association for Educational Assessment, 16th International Conference on Issues in Public Examinations, Maastricht, The Netherlands, June 18-22, 1990.

⁷³*Ibid.*, p. 4.

⁷⁴Kreeft, *Op. cit.*, footnote 71, p. 16.

⁷⁵Eckstein and Noah, *op. cit.*, footnote 6, p. 312.

Current practice has been moving to reduce the uniformity and increase variety and options. Since 1950, the French have changed the *Bac* radically in order to meet demands for a more relevant set of curricula and to open access to a larger group of students. While in the period before 1950 there were 4 options, the *Bac* has diversified into some 53 options and 3 types of *Bac* diploma: secondary (general) education diploma, with 8 options; technician/vocational *Bac*, with 20 options; and, since 1985, a new vocational diploma with 25 options.⁷⁶ The vocational and technical programs have been strengthened and the numbers of students enrolled are also rising.

Indeed, one of the goals of education reform in France has been to democratize the *Bac*. Between ages 13 and 15, the proportion of children attending schools leading to the *Bac* drops from 95 to 67 percent. Among these, one-half actually passed the *Bac* in 1990, i.e., 38.5 percent of students in the relevant age group were eligible for admission to university.⁷⁷ In 1991, 46 percent of the examinees passed.⁷⁸ This **represents** a dramatic reform to the French pyramidal system: in 1955, only about 5.5 percent of French students qualified for university-level education.⁷⁹ The French Government has set a goal for the year 2000 to have 80 percent of students in the age group reach the *Bac* level.⁸⁰ Part of this process is the creation of a number of new technological, vocational, and professional *Bacs*, and better counseling for students concerning specialties, along with restructuring of the *Bac* to make all tracks as prestigious as the “*Bac C*,” the mathematically oriented track.⁸¹

Despite these changes, the *Bac* remains a revered institution in France. It is debated each year as questions and model answers are printed in newspapers after the examinations are given each spring. A central core of general education subjects (e.g., French literature, philosophy, history, and geogra-

phy) is required of all candidates, but different weights are given in scoring them depending on the student's specialization. Examination formats are generally composed of four types of questions: the dissertation—an examination that consists of a question to be answered in the form of an essay; a commentary on documents; open-ended questions; and multiple-choice questions for modern foreign languages.⁸² While MOE formulates the various *Bac* examinations, working from questions proposed each year by committees made up of *lycee* and university teachers, each academy provides its own version from centrally approved lists. Thus questions for each subject, though all of the same nature and level of difficulty, vary from one region to another. Teachers are given some latitude to set their own standards of grading, and there have been concerns regarding a lack of common standards and comparability in the various forms of the *Bac*.

Today the *Bac* can no longer be described as a single nationally comparable examination administered to all candidates. While success in the *Bac* remains the passport to university study, it has been suggested that today there is more than one class of travel in a two-speed university system.⁸³ Thus entry to the slower track remains automatic with the *Bac*, but entry into more remunerative and prestigious lines of study (*classes preparatoires* of *grandes ecoles* and faculties of medicine, dentistry, and some science departments) require high scores in a more difficult *Bac series*. Students who wish to seek admission to the highly selective *grandes ecoles*, which provide superior study conditions and enhanced career opportunities for higher ranks of government service, professions, and business, compete in another examination, the *concours*, usually taken after another year or two of intense preparation. This competition is rigorous; only 10 percent of the age cohort attends the *grandes ecoles*.⁸⁴ Thus, competition to enter a prestigious university or

⁷⁶Sylvie Auvillain, cultural service, French Embassy, Washington DC, personal communication, August 1991.

⁷⁷Embassy of France, Cultural Service, *Organisation of the French Educational System Leading to the French Baccalaureate* (Washington DC: January 1991).

⁷⁸Auvillain, op. cit., footnote 76.

⁷⁹Eckstein and Noah, op. cit., footnote 6, p. 304.

⁸⁰National Endowment for the Humanities, op. cit., footnote 37, p. 9.

⁸¹Ibid.

⁸²Kreeft, op. cit., footnote 71, p. 16.

⁸³Eckstein and Noah, op. cit., footnote 6, p. 3@.

⁸⁴Ibid., p. 304.

professional track has maintained the high value placed on examinations in France.

Germany



Germany is credited with pioneering the use of examinations in Europe. In 1748, candidates for the Prussian civil service were required to take an examination. Later, as a university education became a prerequisite for government service, the

Abitur examination was introduced in 1788 as a means for determining completion of middle school and consequent eligibility for a university entrance.⁸⁵

Today tracking into one of three lines of schooling begins at approximately age 10 in Germany. After completing 4 years of common schooling (*grundschule*), German students move into one of three lines of schooling. The *hauptschule* (main school) or lower general education extends for 5 years and leads to terminal vocational training at about age 16. The *realschule* or higher general education extends for 6 years and directs students to intermediate positions in occupations. The *gymnasium* is the university track and extends for 9 years. There is also a *gesamtschule*: 6 or 9 years of comprehensive schooling containing all three lines. During each of these levels of schooling there are relatively few examinations until their conclusion. There is a reasonable balance in the number of openings for the next level for each track, and examination pressure is not terribly intense at this level.⁸⁶ Because of a traditionally strong and well-respected vocational track, Germany's dual system means that students have several options available to them. Ironically, the traditional distinctions between these two career paths is becoming somewhat blurred and so, by the same token, is the function of the *Abitur*. Increasing numbers of *Abitur* holders are turning toward apprenticeship or technical training rather than

Size	137,743 square miles, slightly smaller than Montana
Population	77,555,000 (1990)
School enrollment	11.0 million (1988)
Age of compulsory schooling	6 to 16
Number of school days	160 to 170 (varies per State)
Selection points and major examinations	1. Tracking at end of common school (age 10) into three lines of schooling, but not via examination 2. <i>Abitur</i> at end of grade 13 for university entrance, determined by each State (<i>and</i>), with oversight by national government
Curriculum control	Land control

academic careers, changing the function of the examination process.⁸⁷

At the conclusion of grade 13 in the *gymnasium*, students take the *Abitur*, which entitles them to study at their local university or any university in Germany.⁸⁸ The specific content of each *Abitur* is determined by the education ministries in the various *lander* (or States) in Germany, within a general framework established by the national Standing Conference of Ministers of Education and Cultural Affairs. It should be noted that the *Abitur*, like the French *Bac*, has changed over the years as the number of students in *gymnasium* has increased, and greater numbers of *Abitur* holders has meant restrictions on their constitutional right to enroll at a university in a chosen course of study. In 1986, 23.7 percent of the relevant age group held the *Abitur*.⁸⁹

In the past, the *Abitur* required candidates to complete an extraordinarily demanding curriculum, but in recent years the breadth and depth of studies has been reduced as variety and options have added diversification to what was once a relatively uniform examination. Demands made on students have been subject to swings; in 1979, candidates could take selected subjects at lower levels of difficulty, but in the fall of 1987 the Council of Ministers reconsidered these changes and restored some of the older regulations and standards, especially limiting candi-

⁸⁵Cummings, op. cit., footnote 43, p. 90.

⁸⁶*Ibid.*, p. 92.

⁸⁷Eckstein and Noah, op. cit., footnote 6, p. 306.

⁸⁸Quite a high number of students do not study at their local university, but at another elsewhere in Germany. Lack of plains at the local university means that some students have to study at distant universities. Reinhard Wiemer, second secretary, German Embassy, Washington DC, personal communication, August 1991.

⁸⁹National Endowment for the Humanities, op. cit., footnote 37, p. 29.

dates' freedom to select subjects at lower levels of difficulty. Students choose four subjects in which to be examined, across three categories of knowledge: languages, literature, and the arts; social science; and mathematics, natural sciences, and technology. Examinations are strongly school-bound, with much effort placed on tying questions to the training provided by a particular school. Even if questions are provided centrally across a /and, different sets are provided from which teachers may choose. In virtually all *lander*, the assessment of the examination papers takes place entirely within the school, by the students' own teachers. Only Baden-Wurtemberg has a system of coassessment by teachers of other schools.⁹⁰

Examinations always consist of open-ended questions, which usually require essay responses. Some examinations are oral, while others, in subjects such as art, music, and natural sciences, may involve performance or demonstration.⁹¹

Despite the open format of the *Abitur*, there has been more concern with comparability across the various *lander* than across individuals, since schooling is a *land* prerogative. There is a delicate balance between State ownership of examinations and national comparability. As a result, some *lander* regard *Abitur* earned in other *lander* with a certain degree of suspicion, limiting student ease of movement to universities across the country and comparability and transferability of credentials.⁹²

Sweden



location, there has been a national curriculum,

Swedish schooling has always been characterized by a blend of central control of curriculum and decentralized management and assessment. In seeking to offer equivalent education to all students, regardless of social background or geographic

Size	173,731 square miles, slightly larger than California
Population	8,407,000 (1990)
School enrollment	1.2 million (1987)
Age of compulsory schooling	7 to 16
Number of school days	180
Selection points and major examinations	1. After compulsory school (age 16) admission to upper secondary school (gymnasieskolan) by marks, not examinations. 2. University entrance by grades or the Swedish Scholastic Aptitude Test (national tests).
Curriculum control	National, common curriculum with local flexibility

accompanied by detailed earmarking of grants to municipal authorities for the organization and administration of schools. Recent reforms have specified that the national government will indicate goals and guidelines, while municipalities are responsible for the achievement of targets set by the national education authority. Each municipality will receive financial support from the national authority, but without detailed spending regulations.⁹³

Compulsory schooling for Swedish children begins at age 7 and extends through grade nine, to age 16. The elementary school (*Grundskola*) is divided into three levels: lower (1 to 3); middle (4 to 6); and upper (7 to 9). Students remain in common heterogeneous classes throughout the first 9 years, but at the upper school level (grades 7 to 9) they begin to choose from a number of elective courses. There is a common curriculum for all schools at each level; those studying any given subject at the same level follow the same curriculum, have the same number of weekly periods, and use common texts and materials. However, it is understood that within the general framework it is up to the teacher to develop his or her own approach to teaching the subject.⁹⁴

After finishing compulsory schooling at age 16, the great majority of students continue on to the integrated upper secondary school or *gymnasieskolan*. At the upper secondary school, there area variety of

⁹⁰Kreeft, *op. cit.*, footnote 71, p.18.

⁹¹National Endowment for the Humanities, *op. cit.*, footnote 37, p. 29.

⁹²Eckstein and Noah, *op. cit.*, footnote 6, p. 314.

⁹³As of July 1, 1991, the National Board of Education and regional country education committees were abolished and a new central education authority was established. Karin Rydberg, *A Redistribution of Responsibilities in the Swedish School System* (Stockholm, Sweden: The Swedish National Board of Education, January 1991).

⁹⁴National Swedish Board of Education, "Assessment in Swedish Schools," informational document, February 1985, p. 1.

courses of study in 2-, 3-, and 4-year programs. Overall some 25 options or lines of study are available, each characterized by a combination of special subjects and a common core of compulsory subjects.⁹⁵ Admission to the integrated upper secondary school is based on teacher grades (referred to as marks) obtained in elementary school, with a certain minimum average required. All subjects (including music, drawing, and handicraft) are included in computing the marks, with none weighted more heavily than any other. In 1983, approximately 85 percent of the age cohort were admitted to the *gymnasieskolan*, with 10 percent applying and not admitted, and about 5 percent not applying to upper level schooling.⁹⁶

Assessment in Swedish schools consists of both marks and standardized tests (*centralaprov*). The individual teacher is solely responsible for the marking, and no educational or legal authority can alter a given mark or force a teacher to do so. Marks are given at the end of each course as a means of providing information to the students and parents on the student's level of success in a course, and are the basis of selection of students for admission to the upper secondary school and to the university. Thus there is considerable effort to provide assurance that marks have the same value, despite the fact that marks are given by thousands of individual teachers across the country.

The main purpose of standardized achievement testing in Sweden is to enable the teachers to compare the performance of their own class with that of the total population and adjust their marking scale. While the *centralaprov* are developed by the national education authority, the tests correspond closely to the syllabi and are aimed at measuring achievement based on national standards. All standardized tests, which are short answer, fill in the blank, and short essay examinations, are centrally developed but administered and graded by the classroom teacher. Detailed instructions on scoring principles are issued by the national board. A sample of results representative of the total population of students tested is submitted to the national board, and marking norms are developed so that test results can be converted into one of the marks on the 5-point Swedish scale. These norms are then sent to all

schools, and teachers mark their tests based accordingly.

Although some tests are used for diagnosis at the classroom level, neither these nor *centralaprov* are used for selection or school accountability in the sense of ranking schools. A large number of standardized tests measuring skills and knowledge are used, along with diagnostic materials. Achievement testing is not conducted until grade eight (in English) and grade nine in Swedish and mathematics. All standardized tests at the elementary level are voluntary for the school and/or teacher; however, about 80 percent of all teachers use them. These tests are used repeatedly over a period of some years and are kept confidential.⁹⁷

In the upper secondary school, the standardized achievement tests must be given in each subject. These, too, have been developed by the national board and are scored by teachers.

Final assessment of each student at the end of a term is a carefully orchestrated business. Teachers keep records of each student's performance on compulsory written tests (in addition to the standardized tests); these are filed and made available when the inspectors from the county education committees visit schools. On these visits, they check to see if the marking principles applied by the teacher are more lenient or severe than national norms. At the end of a term, the teacher surveys all evaluation data collected above (written tests, standardized tests, and observations based on running records) and ranks the pupils in the class from top to bottom on the same 5-point scale.

Here again the standardized tests play an important role. First the teacher calculates the mean of the preliminary marks and records their distribution over the 5-point scale, then compares these data with the mean and distribution of marks obtained by the class in taking the standardized tests. These results are compared and the teacher adjusts the preliminary marks as he or she sees fit, depending on the circumstances surrounding the standardized test (the class may not have covered some part of the standardized test, or there may have been several of the best or the weakest pupils missing when the test was administered, thus skewing results.) The final

⁹⁵*Ibid.*, pp. 3-5.

⁹⁶*Ibid.*, p. 3.

⁹⁷*Ibid.*, p. 13.

judgment is the teacher's, although a meeting called the class conference, attended by the head, assistant head, and all teachers teaching the class for one or more subjects, is also held. At this meeting, comparisons are made between the standard achieved in different subjects and between the achievements of different classes in the same subject. "A teacher who wants to retain noticeable differences between test results and preliminary marks has to convince the class conference that there is a valid reason for doing so."⁹⁸

Sweden abolished its school-leaving examination (for graduation) in the mid-1960s. From that point on, admission to universities and colleges for students coming directly from the upper secondary school has been based entirely on the marks given by teachers. Applicants 25 years or older and with more than 4 years of work experience were admitted based on the Swedish Scholastic Aptitude Test (SWESAT). This test consists of 6 subtests, for a total of 144 multiple-choice items, with a testing time of approximately 4 hours. The SWESAT is administered by the National Swedish Board of Universities and Colleges, with test construction placed in the hands of the department of education at Umea University. About 10,000 persons take the test each year. The selection procedure was part of an elaborate system of quota groups to ensure a fair distribution of openings for different groups of applicants. There are three groups: those submitting formal measures of academic ability-grades and SWESAT for those who have not completed upper secondary education; those relying on work experience-which for all groups of applicants may compensate for a low score on academic ability; and a small number of places for those accepted for special reasons, despite low scores.

In the 1970s and 1980s, the number of applicants to higher education greatly exceeded the number of available places, and this created debate. The existing system of quotas was criticized for being cumbersome, uniform, and complex. Furthermore, the use of work experience was criticized on the grounds that it delays the transition to higher

education. In fact work experience has become almost compulsory for many programs in high demand. (Today the average age of a first-year freshman in Sweden is 23.) The fact that practically all experience is given credit, regardless of relevance to the study program in question, has also been debated. Some believe the system should give weight largely to academic ability as a better predictor of success in higher education.

As a result of this debate, the Swedish Parliament established a new scheme for selection to higher education that more strongly stresses the need for measures of academic ability and restricts the role of work experience. The new system, which went into effect in July 1991, uses several factors for determining admission. Average grades from upper secondary school will continue to constitute a major factor in the selection process. (Between one-half and two-thirds of all students will be selected on the basis of grades alone.) A general aptitude test (currently the SWESAT) is open to students leaving upper secondary school as well. This is seen as an alternative path to higher education for those who do not have sufficient grades. Between one-third and two-thirds of all students will be selected on the basis of the test results. Finally, flexibility is being added to ensure that a small number of students can be admitted on an individual basis.⁹⁹ It is not yet clear what the impact of these changes will have on school curriculum across Sweden.

England and Wales



The Education Reform Act (ERA) of 1988 set in motion a major overhaul of the education system of the United Kingdom (England, Wales, and Northern Ireland).¹⁰⁰ Although authority over the schools had been shifting from local to central government at least since the second World War, the 1988 reforms were seen by many as a watershed event. One analysis by comparative education re-

⁹⁸Ibid., p. 17.

⁹⁹Hans Jansson, "Swedish Admissions Policy on the Road From Uniformity and Central Planning to Flexibility and Local Influence?" paper prepared for the International Association of Educational Assessment, November 1989. See also Ingemar Wedman, Department of Education, University of Umea, Sweden, "The Swedish Scholastic Aptitude Test: Development, Use and Research" unpublished document, October 1990.

¹⁰⁰There are actually three education systems in the United Kingdom: one for England and Wales, a second for Scotland, and a third for Northern Ireland. This report deals predominantly with England and Wales, but all three systems are reforming curriculum and assessment programs.

Size	94,226 square miles, slightly smaller than Oregon
Population	57,121,000 (1990)
School enrollment	10,089,000 (1983)
Age of compulsory schooling	5 to 16
Number of school days	192 ^a
Selection points and major examinations	<ol style="list-style-type: none"> 1. New national assessments at age 7, 11, 14, 16 (not for selection) 2. Two-tiered school-leaving examinations: General Certificate of Secondary Education at age 16 or earlier; "A levels" at grades 11 or 12 (sixth form) at age 18 (all set by local boards, national oversight, considered for university entrance)
Curriculum control	National, central control (since 1988)

^aWayne Riddle, Congressional Research Service, personal communication, Nov. 26, 1991.

searchers concluded that the reforms “. . . represented an abrupt acceleration of the otherwise glacially slow process of transferring authority over the schools from local to central government.”¹⁰¹

England always had a diverse and decentralized school system. The great universities and “public” schools,¹⁰² which were closely tied to the Church of England, existed for the upper classes; there was no need for selective entrance examinations, given that student qualifications were not an issue for admissions.¹⁰³ In the middle of the 19th century, England’s highly decentralized system distinguished it from other European countries, which already had strong central curricula and uniform school-leaving examinations. To bring some order to the system, the British Government instituted the “payment by results” system. Beginning in 1861, local governments whose students performed well on a special national test received extra subsidies. The goal of this policy was to promote quality in key subject areas. There was no attempt to create a central curriculum.¹⁰⁴ This testing program was eventually scuttled because of dissatisfaction with the inequalities it aggravated. Schools that had the most difficult

problems were those that suffered most under the system; essentially the rich got richer.

Following World War II, in an effort to democratize secondary school selection procedures, the 11+ examination was developed. These were local examinations, run by local education authorities (LEAs). The goal was to track students at age 11, according to ability, as measured on the examination and according to need. Roughly 20 percent of students were tracked into grammar school (i.e., the college preparatory track) and the rest into secondary “modern” schools. As LEAs introduced comprehensive schools in place of the grammar and secondary moderns, the 11+ was no longer needed. Although it is still in use in a small number of places in England and Wales, by and large the 11+ was dropped during the 1960s and 1970s.

The General Certificate of Secondary Education (GCSE) continues the tradition of local control of curriculum and testing. Although the concept of merging the prior “ordinary” examination (“O levels”) and the GCSE examinations goes back to the early 1970s, the first GCSE examinations were administered in 1988. The GCSE became the single examination, mirroring the switch from the grammar and secondary moderns to one comprehensive school. The GCSE is taken by students at the age of 16 or earlier. Local groups of teachers and school administrators, through the examining boards, introduce examination topics related to their own syllabi. A central School Examinations and Assessment Council, established by Parliament, establishes national examination criteria to which all GCSE syllabi and examinations must conform. Recruitment into certain jobs and selection into advanced training are influenced by the number and quality of passing grades on the GCSE.

More advanced examinations, the ‘A levels’ are also offered in the upper grades of comprehensive school (age 18). Success on at least three A levels has become an important criterion for advancement to university study. Thus the school-leaving examination system in the United Kingdom has evolved into a two-tiered examination system. A recent

¹⁰¹Noah and Eckstein, op. cit., footnote 44, p. 25. This characterization may be somewhat overstated, given that local management of schools remains an important component of the school system. Robert Ratcliffe, academic programs officer, The British Council, personal communication, Aug. 15, 1991.

¹⁰²English “public” schools would be called “private” in the American idiom.

¹⁰³Cummings, op. cit., footnote 43.

¹⁰⁴Ibid., p. 93.

survey of 16-year-olds in England showed slightly over one-half planning to continue their education. About one-third of the country's 16-year-olds, who achieved grades of A, B, or C (on a scale of A to G) on five or more of their GCSE examinations, are most likely to continue. In 1988-89, 22 percent of all 18-year-olds in England passed one or more A-level examinations; 12 percent, three or more.¹⁰⁵ Students have, in the past, been able to select their own subjects for the GCSE and its predecessors and for the A levels.¹⁰⁶ There is some concern that early specialization in grades 11 and 12, to prepare for A levels, is one factor causing many students to abandon study in mathematics and the sciences at age 16 in favor of the humanities or social sciences.

The background for the 1988 reform was similar to the push for educational changes in the United States: business people were complaining that students arrived at the workplace lacking basic skills, while others were troubled by inequalities in teaching, resources, and by an education system out of sync with technology. The Conservative Government under Margaret Thatcher put into place a reform bill that forced the issue. As the chief executive of the newly established National Curriculum Council noted: "The educational establishment, left to its own, will take a hundred years to buy a new stick of chalk. . . . In the end, to say: 'It's time you guys got on with it; here's an act and a crisp timetable' was probably necessary." ¹⁰⁷

First and foremost, the ERA defined a comprehensive national curriculum for all public school students ages 5 to 16. These students are to take foundation subjects: core subjects are English (Welsh, in Wales), mathematics, and science, plus, for 11- to 16-year-olds, technology (including design), history, geography, music, art, physical education, and modern foreign language. Attainment targets set general objectives and standards for 10 levels covering the full range of pupils of different abilities in compulsory education. Average pupils will reach level two by age 7; each new level represents, on average, 2 years of progress. The statements of attainment provide the basis for the assessment

arrangements. Assessment is to take place by classroom teachers throughout the year, with special soundings via national tests known as standard assessment tasks (SATs) given at or near the completion of each of four 'key stages' of teaching (ages 7, 11, 14, 16).

The assessments are meant to serve multiple purposes:

... **formative**, providing information teachers can use in deciding how a pupil's learning should be taken forward, and in giving the pupils themselves clear and understandable targets and feedback about their achievements; **summative**, providing overall evidence of the achievements of a pupil and of what he or she knows, understands and can do; **evaluative**, providing comparative aggregated information about pupils' achievements as an indicator of where there needs to be further effort, resources, changes in the curriculum; and **informative**, helping communication with parents about how their child is doing and with governing bodies, LEAs and the wider community about the achievements of a school.¹⁰⁸

The objective is to keep the schools working within a national framework but with local discretion in implementing the curriculum. As parents can now send children to any school they choose, it is anticipated that parents will compare published examination results of schools, and thus schools will try to raise standards to attract more pupils.¹⁰⁹ But there is concern that comparisons may mask differing social and economic levels of students, and that problems associated with the "payment by results" approach of 100 years ago could return. Teachers also feel overwhelmed by the requirements of the program: the double system of assessment at key stages—with the SATs as well as continuous assessment in the classroom—means that British school children will soon be the most assessed in Europe.

The program is being implemented at the primary level in the spring of 1991 and will be phased in over the next 3 years. Secondary students may be assessed through GCSEs or according to National Curriculum assessments at age 16. GCSE criteria

¹⁰⁵National Endowment for the Humanities, op. Cit., footnote 37, p. 45.

¹⁰⁶Few schools allowed students to omit *mathematics* and English for the General Certificate of Secondary Education and its predecessors, but roles about what must be studied at this level will become tighter under the national curriculum assessment. Nuttal, op. cit., footnote 14.

¹⁰⁷Tim Brookes, "A Lesson to Us All," *The Atlantic*, vol. 267, No. 5, May 1991, p. 28.

¹⁰⁸Department of Education and Science, *National Curriculum: From Policy to Practice* (Stanmore, England: 1989), p. 6.

¹⁰⁹Brookes, op. cit., footnote 107.

and syllabi will be brought into line with the statutory requirements for attainment targets, programs of study, and assessment strategies, but the relationship between National Curriculum's 10 levels of attainment and the GCSE grades has yet to be determined.¹¹⁰ In early 1991, plans were announced to require all students to take GCSEs in the three core subjects of English (or Welsh), mathematics, and science. The study of either history or geography, technology, and a modern foreign language is also compulsory to age 16. Students can choose whether to have their competence in these and other subjects assessed by GCSE examinations.¹¹¹

The SATs are one of the most interesting features of the program, and the feature most likely to influence curriculum. As in most European testing programs, the SATs have only open-ended questions. Many innovative testing approaches were developed for an earlier comprehensive assessment England embarked on in 1975.¹¹² These innovative test items and formats are the basis for many of the performance testing items that are to become the backbone of the SATs and classroom assessment procedures under the new program.

A nationally representative sample of students at ages 11, 13, and 15 were tested in a survey similar to NAEP. The 1975 goal was to assess the achieve-

ment and knowledge of student performance in four areas: mathematics, language, science, and foreign languages.

Mathematical abilities were tested in several formats, including 50 short-response items drawn from a total of 700 test items in each survey. A subsample of students in each age group were given written tests of problem-solving skills; another subsample of 1,200 students in each age group were given oral tests of problem-solving tasks. The mother language survey assessed reading, writing, and "oracy," a term coined for its analogy to literacy as a measure of the ability to communicate effectively in a spoken as opposed to written medium. The science assessments were made up of individual and small group tasks emphasizing practical skills performed at a number of "stations." Foreign language testing used oral and written testing formats.

The program led to the evolution and application of innovative techniques to assess student performance, such as mathematical skills in a practical context, especially those whose mathematical abilities were masked by reading difficulties; written and spoken skills in the mother tongue and in foreign languages; and practical assessments in science.

¹¹⁰Department of Education and Science, op. cit., footnote 108, paragraph 6.7.

¹¹¹National Endowment for the Humanities, op. cit., footnote 37, p. 45.

¹¹²Clare Burstall, "Alternative Forms of Assessment: A United Kingdom Perspective," *Educational Measurement: Issues and Practice*, vol. 5, No. 1, spring 1986.