This article describes the education quality control systems (for mathematics) used by those countries that performed best on the Third International Mathematics and Science Study (TIMSS). Enforced quality control measures are defined as "decision points"—where adherence to the curriculum and instruction system can be reinforced. Most decision points involve stakes for the student, teacher, or school. They involve potential consequences for failure to adhere to the system and to follow the program at a reasonable pace. Generally, countries with more decision points perform better on the TIMSS. When the number of decision points and TIMSS test scores are adjusted for country wealth, the relationship between the degree of (enforced) quality control and student achievement appears to be positive and exponential. Conclusion: The more (enforced) quality control measures employed in an education system, the greater is students' academic achievement.

BENCHMARKING TO THE WORLD'S BEST IN MATHEMATICS

Quality Control in Curriculum and Instruction Among the Top Performers in the TIMSS

RICHARD P. PHELPS
Westat

We have made considerable progress because we resisted the temptation to put our faith in any single gimmick or formula for school improvement. School systems are complex—and looking for a simple solution is, well, simple-minded.

—Rod Paige, as Superintendent of the Houston Independent School District (currently, he is U.S. Secretary of Education)

Integrated systems that work well together are the essence of civilization.

—Irving Wladawsky-Berger, general manager, Internet Division, IBM

AUTHOR'S NOTE: *The author would like to acknowledge the contribution of others to this work. Lois Peak of the U.S. Department of Education's Planning and Evaluation Service conceived the data collection, helped design the data collection instrument, and recruited the country expert respondents. She is not responsible, however, for the analysis and, therefore, should*

EVALUATION REVIEW, Vol. 25 No. 4, August 2001 391-439

The United States has participated in five international assessments of student achievement in mathematics and science since the 1960s. Each time, the comparison of U.S. student performance to their international counterparts' has provoked widespread interest from researchers, policy makers, and the public at large. The occasions have prompted wholesale critiques and defenses of the U.S. education system in the popular press. The scholarly press, in the meantime, has been filled with studies of U.S. relative achievement in the context of various background factors, such as the average educational attainment level or socioeconomic status of the test-takers' parents or the level of public education funding.

Most attention has focused on the validity of country-average test score comparisons in the light of differences in the mechanics of test administrations and sample selection across countries, with critics claiming that the differences nullify valid comparisons. Defenders of the country-average test score comparisons have argued that the differences in the test administration mechanics do not invalidate comparisons because they are not large enough or they should average out over time. They argue that comparative U.S. mathematics performance at the 8th-grade level has been relatively consistent over five assessments and three decades.

The background analyses probing the deepest have searched for explanations of relative achievement in the curriculum of each country. The Second International Mathematics and Science Study (SIMSS) in the early 1980s spawned *The Underachieving Curriculum*, a critique of the prevailing U.S. mathematics curriculum written by some of the U.S. researchers directly involved in building and analyzing the SIMSS database (McKnight et al. 1987). Some of the same researchers were involved in building and analyzing the database for the Third International Mathematics and Science Study (TIMSS), administered in the 1994-1995 school year. Their main curriculum analysis studies, *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education*, *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Mathematics*, and *Characterizing Pedagogical Flow: An Investigation of Mathematics and Science Teaching*, echoed the critical refrain of *Underachieving Curriculum* (Schmidt et al. 1996a, 1996b, 1997). The U.S. mathematics curriculum, by comparison with its international counterparts, lacked focus and depth. One of the most widely quoted phrases from one of the study's authors characterized the U.S. math curriculum as "a mile wide and an inch deep."

---

not be held liable for any of its errors. Ellen Pechman and Rolf Blank reviewed early drafts of the questionnaires used in this study and provided helpful comments on them. The author retains all responsibility for any errors in this article.

Other studies have looked deeply at instructional practices across countries. Over the past two decades, Harold Stevenson and James Stigler (1992) have observed and compared classroom culture and instructional practices in the United States and East Asian countries and have discovered some highly enlightening contrasts. Coincident with the TIMSS, George Stigler videotaped many hours of secondary-level mathematics classroom instruction in samples of German, Japanese, and U.S. schools. The contrasts in instructional style, demeanor, and content are striking (Office of Educational Research and Improvement 1997a).

Still other studies have looked more explicitly at the benefits, methods, and feasibility of benchmarking curricular and instructional practices across countries. To this effort, some researchers have focused on content standards (Beatty 1997; Resnick, Nolan, and Resnick 1995; Nolan 1997; Louis and Versloot 1996) and others on performance standards (Britton and Raizen 1996; Eckstein and Noah 1993; Gandal 1997; Stevenson and Lee 1997). Still other researchers have argued for more comprehensive comparisons of education systems across countries and the impact of many systemic influences on curriculum and instruction (Bishop 1997; Mullis 1997a), or they have advocated efforts toward benchmarking entire systems of curriculum and instruction (Cross and Stempel 1995; Shanker 1996; U.S. Department of Education 1995).

This report aims to supplement the aforementioned curriculum and instruction studies with a look behind the scenes at the formation and implementation of both. It takes one giant step back in the process to better understand the superstructure of other countries' curriculum and instruction systems and the "glue" that holds that superstructure together, to better understand how other countries see to it that the curriculum they intend is attained. Essentially, it focuses on how top-performing countries control quality in their curriculum and instruction systems.

This article exploits information gathered in study sponsored by the U.S. Department of Education's Office of Educational Research and Improvement and National Center for Education Statistics, and a variety of other sources, in an attempt to capitalize on the occasion and the wealth of information provided by the TIMSS, to better understand U.S. mathematics and science education in its international context.

THE TIMSS

There are perhaps no singular events that elicit more public judgment of the quality of U.S. elementary and secondary education than the periodic release of results from international student assessments. The TIMSS,

administered in 1994-1995, was the largest such assessment ever, with more than 40 countries participating at one or more of three grade levels—the rough equivalents of our 4th, 8th, and 12th grades. Results for the grade level at which the most countries participated—8th grade—were released first.

When the mathematics performance of U.S. 8th graders was compared to their international counterparts' in the summer of 1996, it seemed to reaffirm in the minds of many U.S. observers the legacy of pessimism from earlier international assessments. Among the 40 countries with student scores meeting minimal statistical requirements for comparison, U.S. 8th graders scored lower than 8th graders in 20 other countries and higher than those in only 7, when measured by a multiple comparison procedure involving all participating countries. U.S. students' scores were on a par with those of students in 13 remaining countries (Beaton 1996, 23).

The performance of U.S. 4th graders, made public the following summer, seemed much better. A multiple comparison procedure showed U.S. 4th graders scoring below their counterparts in 7 countries, above those in 12, and on a par with those in 6 other countries (Mullis 1997b, 25).

In between the relatively strong U.S. 4th-grade performance and the relatively weak U.S. 8th-grade performance were three grade levels and a steep decline in U.S. relative performance. Among all the 25 countries that participated at both the 4th- and 8th-grade levels and met minimal statistical requirements for comparison, the "synthetic gain" in mathematics achievement between the 4th and 8th grades appeared to be the smallest in the United States (Mullis 1997b, 43). One could speculate that the longer students stayed in U.S. schools, the less they learned, by comparison with average academic progress in other education systems.

The release of the 12th-grade results in 1998 only seemed to confirm the most pessimistic predictions. The unfortunate trend in relative U.S. student performance continued downward through the upper secondary years (Mullis et al. 1998).

EXPLANATIONS FOR THE U.S. TEST PERFORMANCE

Ultimately, however, test score comparisons alone do not tell the whole story. There can, after all, be many explanations for any country's disappointing test performance. An explanation might lay in the mechanics of the test administration, perhaps, if one country's students were younger in age for their grade level, or the test was given earlier in the academic year. Likewise, an explanation might lay in the social background from which each student

emerges if one country has relatively higher proportions of nonnative speakers of the primary language or households in poverty, for example.

Likewise, some explanation might lay in the structure and procedures of each country's education system. The aforementioned reports from the U.S. TIMSS Committee argued that the U.S. mathematics curriculum lacks the focus and depth often found in other countries. One could argue that the videotape studies of George Stigler showed the same to be true in the conduct of classroom mathematics instruction.

Ina Mullis, of the International TIMSS Center at Boston College, observed that the top-performing countries at the 8th-grade level were more likely to have high-stakes examination systems than were other countries (Mullis 1997a). John Bishop, the Cornell labor economist, has found statistically significant effects from the existence of high-stakes examination systems on student test performance using data sets of the 1991 International Assessment of Educational Progress (IAEP) across countries or across Canadian provinces, of the Scholastic Assessment Test across U.S. states, and now across countries with the TIMSS. His discovery of significant effects is all the more remarkable because the high-stakes tests in some of the countries, states, or provinces are upper secondary-level exit examinations, given to students when they are 17, 18, or 19 years old, whereas the tests providing his measures of achievement in the case of the IAEP and the TIMSS were administered to 13-year-olds. He calls the alleged effect of high-stakes upper secondary exit exams on the behavior of students and teachers at the lower secondary level a "backwash" effect (Bishop 1997).

In another study, "Impacts of School Organization and Signaling on Incentives to Learn in France, England, Scotland, the Netherlands, and the United States," Bishop (1993) expanded his analysis to include "signals" of student performance and expectations other than those derived from examinations, such as the publication of exam results, retention in grade, selection of students for different curricular tracks (e.g., academic, vocational, general), amount of homework required, "looping" of teachers over several grade levels with same students so that the person responsible for teaching particular students was identifiable, and so on.

BENCHMARKING

Coincident with the student performance comparisons of the past decade, several groups have studied the curricula of other countries and compared them with curricula typically found in the United States. Most commonly,

these studies have focused on the content of mandated, large-scale examinations as the most concise representations of a curriculum. Under Secretary Lynn Cheney in 1991, the National Endowment for the Humanities translated and published side-by-side comparisons of secondary-level history examinations from France, Germany, Japan, England and Wales, and Belgium (National Endowment 1991). The New Standards Project (1994) and the National Center on Education and the Economy (1994) translated and compared several countries' mathematics examinations. The National Center for Improving Science Education translated and compared several countries' science examinations (Britton and Raizen 1996). The American Federation of Teachers has done the same in several subject areas (e.g., American Federation of Teachers 1995a, 1995b). The National Center for Education Statistics sponsored work by the Pelavin Research Institute (1996) comparing national assessments in Canada, England and Wales, France, and the United States.

The Council for Basic Education has gone a step further in its Schools Around the World Project, enlisting the cooperation of classrooms in eight countries to participate in an exercise that will compare several kinds of student work, including homework and term papers, rather than just examinations (Council for Basic Education 1996).

The American Federation of Teachers has proposed institutionalizing efforts such as these while providing an ongoing reference source for U.S. schools in a U.S. national benchmarking institute. The institute would assist U.S. states and local school districts to conduct systematic exercises in benchmarking elements of their curriculum and instruction to those in other countries, states, and districts (American Federation of Teachers 1995a, 1995b).

All these groups have searched for appropriate benchmarks for help in designing U.S. curriculum and instruction to appropriate levels of depth and difficulty. All these groups realize, however, that benchmarking simply to a result does little, in and of itself, to help achieve the result. To use benchmarking to achieve a desired result, one must benchmark to a behavior that one believes will produce the result.

RESEARCH FOCUS AND SURVEY

This study attempts to understand the superstructure of the education systems that support curriculum and instruction leading to high performance. What is the glue that holds that superstructure together? Given the "intended curriculum" in each country, how is the intended curriculum implemented and attained? How do top-performing countries control quality in their curriculum and instruction systems?

In 1997, a detailed, 15-plus page questionnaire on this topic was assembled, and knowledgeable experts in their respective countries' education systems were asked to fill out and return them. The questionnaire, with abbreviated versions of each country's responses, is available from the author upon request.

The long, but accurate, title of the survey was "Exploratory Survey on the Relationships Among Content Standards, Textbooks, Student Performance Standards, and Examinations in Secondary School Mathematics." The title emphasizes the interest in the connections between the main elements of any country's curriculum and instruction system. The intent of the survey was to learn how and to what degree these elements were integrated in top-performing countries.

Given limited resources, the survey focused on mathematics alone. The reader should realize that conclusions drawn from studying curriculum and instruction in one subject area are not necessarily wholly applicable to others.

The questionnaire consisted of two parts. Part 1 contained questions pertaining to content standards, textbooks, student performance standards, and international benchmarking activities. Part 2 focused on the application of student performance standards at decision points. Experts filled in a separate Part 2 for every decision point their country used. A *decision point* was defined as "an occasion when a student performance standard is actually applied: a judgment is made—for example, that a student achieves or does not achieve a standard—and an appropriate consequence results." Most often, decision points consist of high-stakes tests or selective admissions to certain schools or curricular tracks.

SELECTION OF FOCUS GROUP OF COUNTRIES

Countries from which the United States could learn something on the topic of education system integration and quality control were selected. The size of the group was limited to nine.[1] The first criterion for selection was a superior performance on the TIMSS 8th-grade mathematics test.

Rather than just pick the nine countries ranked highest by average 8th-grade TIMSS mathematics score, however, other criteria were imposed on the selection. It was deemed important, for example, to make sure that some countries with some basic education system characteristics similar to our own, such as large size and a federal structure, were picked. Singapore's students scored higher than any other country's on 8th-grade mathematics, but even though we might be able to learn a lot from Singapore's education system, the United States cannot become very much like Singapore. Singapore is of relatively small size and has a highly centralized education system,

TABLE 1: Focus Group of Countries Ranked in Order of Average Eighth-Grade Third International Mathematics and Science Study Mathematics Score, by Reason for Selection

Country	*Education System That Is Not Centrally Controlled*	*Large*	*Diverse Population*	*Given the Character of Countries Above Already Picked, the Addition of This Country Offers Diversity in Its Geographic Location or Governance Structure*
Singapore[a]			Yes	
Korea[a]				Yes
Japan[a]		Yes		Yes
Hong Kong				
Belgium[a], Flemish			Yes	Yes
Czech Republic[a]				Yes
Slovak Republic				
Switzerland[a]	Yes		Yes	Yes
The Netherlands[a]	Yes		Yes	Yes
Slovenia				
Bulgaria				
Austria				
France[a]		Yes	Yes	Yes
Hungary				
Russia	(changing)	Yes	Yes	
Australia[a]	Yes	Yes	Yes	Yes
Ireland				

a. Country selected for our focus group.

both natural advantages for creating cohesion. So, although Singapore was included in the focus group, there was felt to be no need to include more countries like Singapore but some need to include countries more like the United States (i.e., large, diverse, with federal system).

So, moving down the list of top-performing countries, selection was biased in favor of countries that could both diversify the focus group and ensure that some countries "more like us" were included. Thus, Australia was selected over Austria and Hungary, for example, because of its large size and federal structure and because other countries like Austria and Hungary had already been selected.

Table 1 lists the 13 countries with the highest average 8th-grade TIMSS mathematics scores and other criteria by which countries were selected for inclusion in the focus group.

PROGRESS OF THE SURVEY

In time, some very detailed, thoughtful responses were returned; other brief, but still very thoughtful, responses were returned; and two countries, Australia and the Netherlands, did not respond. Survey results were then supplemented with information from other sources.

Responses were received from experts in Singapore, Korea, France, Japan, Switzerland, the Czech Republic, and the Flemish Community of Belgium. Other sources were consulted to learn about the Netherlands because it provides such an interesting contrast to Flemish Belgium and shares so many important governance characteristics with the United States. Not enough information was gathered to provide a representative picture of Australia, unfortunately, and it had to be dropped from the group.

All countries that returned questionnaires provided fairly complete and thoughtful responses to Part 1, which posed questions on standards, textbooks, and benchmarking, with the exception of Section C on student performance standards. Part 2, which posed questions regarding the application of student performance standards at decision points, received a fairly poor response. One cannot be certain of the reason, but some respondents may not have well understood what was meant by "student performance standard." Fortunately, some country experts provided equivalent information in their other responses to Part 1. Information provided in the questionnaires was verified by country experts in the United States or from written sources.

For the remaining countries, and to fill in any missing information from the responding countries, other sources of information were sought. These other sources are listed by country in the appendix.

In the end, the exploratory survey provided results that traced the outline of the curriculum and instruction picture, but, ultimately, no information from the survey alone was used to draw any conclusions in this analysis.

ANALYSIS: HOW COUNTRIES CONTROL QUALITY IN CURRICULUM AND INSTRUCTION

COHERENCE

The analysis adopts the common and useful framework of vertical and horizontal coherence, widely used by education policy analysts in recent years as a rough device for measuring the degree to which curriculum and instruction systems are integrated. A completely coherent system would be

one with a seamless integration among the various system elements: content standards (the "intended curriculum") represented completely and precisely in textbooks, student performance standards, and examinations, and evaluations of performance representing completely and precisely the mastery of the content.

A system with complete vertical coherence is one in which the intentions of educators at the top of the system (e.g., in the country or state education ministry) are represented completely and precisely in the classroom. A system with complete horizontal coherence is one in which the content standards are represented completely and in precisely the same way in every classroom throughout the country or state.

No country- or state-level education system can have complete, absolute coherence in curriculum and instruction, of course. Only a system consisting of a single classroom with a single teacher who also serves as education minister could offer that. But some education systems make a greater effort than others to maintain coherence, and some are more successful than others in that effort.

Of course, maintaining coherence may be easier in some contexts than in others. Education systems that are small and highly centralized (e.g., Singapore) probably pose the least amount of difficulty. Education systems that are large and highly fragmented among levels of government and types of governance (e.g., United States) probably pose the greatest amount of difficulty. Some might argue, however, that the system of governance in education itself should be considered as a characteristic that can be altered, along with others, if need be, to improve system coherence.

VERTICAL COHERENCE

Vertical coherence implies a process whereby there is a match between the intended curriculum and the "attained curriculum": what students learn. Between the initial writing of content standards and the final mastery by students of subject matter, there may be many interim steps, several layers of government, several organizations involved, a long time lag, and other potential barriers to complete coherence. How does an education system maintain coherence in the face of natural entropy?

Singapore provides a good example of a country with a high degree of vertical coherence. The Ministry of Education (1993) writes content standards, curriculum guides, and some textbooks. Some content is prescribed by the University of Cambridge syndicate, of which Singapore is a member. The ministry trains the teachers in a single, in-house training institute. The ministry

has jurisdiction over all schools, both government and government-dependent private schools. The ministry sends out subject specialist inspectors to monitor classroom instruction. Whenever there are curriculum changes, teachers attend workshops on these changes run by the ministry. Teachers participate in writing and scoring national examinations. There are lots of examinations: on exiting primary school (and getting places in a secondary school of choice), on exiting lower secondary school, on exiting upper secondary school, and for selection to preferred curricular tracks at various points.

In other words, in Singapore, the Ministry of Education (1993) controls most aspects of the process itself, closely monitors classroom instruction, and ties teachers to the examination program by involving them in writing and scoring them.

Korea's system has more variety and diversity in some ways. There are more curricular tracks, particularly for vocational education. Regional governments have some say in how the system is run. Still, the curriculum and instruction process is highly centralized, course content is prescribed by the ministry, and the ministry administers standardized, high-stakes examinations.

Another avenue, outside a single, centralized authority, for maintaining a high degree of vertical coherence is within subject areas rather than over the system as a whole. For example, in some countries, mathematics departments in universities train mathematics teachers, grant teacher certifications, write content and student performance standards, write texts, inspect classes, and write and score examinations with teachers' help. Elements of this kind of vertical coherence exist in the Netherlands and Switzerland.

HORIZONTAL COHERENCE

Horizontal coherence implies a process whereby the curriculum and instruction in one part of a country or state matches that in another part of the country or state. How does a country maintain horizontal coherence? It can mandate a common core curriculum; use common, unique textbooks; train teachers in a single institution or in multiple institutions with one prescribed, standardized program; centralize the approval of curriculum plans, timetables, and inspections; inspect school classrooms with subject area experts to see if curriculum and timetables are followed; establish networks of subject-area professionals and involve them in writing standards, doing inspections, and writing and scoring examinations; and advertise standards to the public so they can hold their local schools accountable.

The Netherlands provides a good example of an education system that maintains a high degree of horizontal coherence. There are few limitations on

forming a school; most any religious or nonreligious organization can do it. Any one school may have no necessary connection with any schools at lower or higher levels of education nor any administrative connection with the central government. Moreover, there are no systemwide content standards or core curriculum. The Netherlands maintains horizontal coherence primarily through frequent administrations of nationally standardized high-stakes examinations.

Flemish Belgium maintains horizontal coherence without standardized tests but with common texts and curriculum guides and widespread public relations efforts that educate the public about what to expect from their local schools.

Table 2 lists various methods that each of the countries in the focus group use to maintain vertical or horizontal coherence. "Yes" means that a country uses the method, "no" means they do not, and a blank cell represents a lack of sufficient information to make a judgment.

Few of the quality control methods implementation listed in Table 2 are prevalent in the United States.

TWO GROUPS OF COUNTRIES

The focus group of countries divides into two natural groups, as characterized by their governance and their methods for maintaining coherence.

Group 1: Highly centralized systems with highly prescribed content and performance standards—Singapore, Korea, Czech Republic, France, Japan.

It is perhaps easy to understand how these countries manage quality control and maintain coherence in curriculum and instruction. Many of the factors involved are controlled centrally. For example, the already-described Singaporean and Korean systems are highly centralized.

France also has highly centralized standard-setting procedures, and all teachers are employees of the central government. There is some variety to examination writing from regional centers and some variety of textbooks. Still, examinations are mostly similar, they are high stakes, and they are numerous and prominent. The Conseil National de Programmes operates much like an Inspector General's office, with inspectors drawn from among the ranks of their office, of secondary school teachers, of university professors, and of Ministry of Education (1993) officials.

In the Czech Republic and other formerly communist Eastern European countries, they are in the process of moving away from this model. There are

TABLE 2: Education System Practices That Produce Vertical or Horizontal Coherence in Curriculum and Instruction, by Country and Practice

Quality Control Practice	*Belgium, Flemish*	*Czech Republic*	*France*	*Korea*	*Japan*	*The Netherlands*	*Singapore*	*Switzerland*
Practices that produce both vertical and horizontal coherence								
Content standards are fixed and are expected to be followed as a core curriculum	Yes	Yes	Yes	Yes	Yes		Yes	Yes[a]
Teachers are required to teach core curriculum	Yes	Yes	No	Yes	Yes		Yes	No
Common or unique textbooks are required to adhere closely to the content standards	Yes	Yes		Yes	Yes		Yes	Yes[a]
Centralized approval of curriculum plans, course timetables, and inspections	Yes	Yes	Yes	Yes	Yes		Yes	Yes
Selective admission to curricular tracks based on standards	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Inspections are done in classrooms, in some cases by curricular experts, and are standards based	Yes	Yes	Yes	Yes		Yes	Yes	
Train teachers in a single institution or in multiple institutions with standardized, prescribed programs						Yes	Yes	
High-stakes exit examinations from lower secondary level are standardized			Yes	Yes	Yes	Yes	Yes	Yes
High-stakes exit examinations from upper secondary level are standardized		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Practices that produce vertical coherence								
Some teachers have the same group of students for more than 1 year	Yes	Yes		Yes	Yes		Yes	
Curricular tracking by school	Yes	Yes	Yes	Yes	Yes	Yes		Yes

(continued)

TABLE 2 Continued

Quality Control Practice	*Belgium, Flemish*	*Czech Republic*	*France*	*Korea*	*Japan*	*The Netherlands*	*Singapore*	*Switzerland*
All students in a school (which may have a curricular focus and be selective) follow the same course of study	Yes	Yes		Yes	Yes			
Establish networks of subject-area professionals and involve them in writing standards, doing inspections, and writing and scoring examinations	Yes		Yes		Yes	Yes		Yes
Employers are directly involved in some aspects of the process	Yes			Yes	Yes	Yes		Yes
Practices that produce horizontal coherence								
Schoolwide curriculum plans with target goals are used to standardize and integrate curriculum and instruction	Yes			Yes		Yes		
Students do not begin homework during class time as instruction time is used to keep a set pace (> 50% of classrooms respond Yes)	Yes	Yes		Yes	Yes		Yes	
Involve educators from around the country in developing the standards	Yes		Yes	Yes	Yes	Yes	Yes	
Involve educators from around the country in writing and revising the textbooks			Yes		Yes			Yes
Advertise common standards to public so they hold local schools accountable	Yes		Yes					
Selective admission criteria to curricular tracks are standardized	Yes		Yes	Yes	Yes	Yes	Yes	Yes

a. Yes for lower secondary, no for upper secondary.

discussions of lowering the required proportion of the core curriculum from 80% to 100% of what is taught to 50% of the curriculum or less, allowing more local control over the curriculum and reducing emphasis on math and science to make room in the curriculum for more social studies and humanities courses. It will be interesting to see if the high performance in math and science holds up in these countries after these changes are made.

Group 2: Decentralized systems with unprescribed aspects to the process of content or performance standard setting—Switzerland, Flemish Belgium, the Netherlands.

Of the focus group countries, Switzerland is closest in its governance structure to the United States but is different in other ways. For example, each Swiss teacher is supervised by an inspector; there are several curricular tracks and all have high-stakes exit examinations (some cantons also have exit exams at three levels: primary, lower secondary, and upper secondary); some of these tracks are also very selective in their entry; the national government does have some say over certification requirements at the upper secondary level; there are several national organizations, such as the Cantonal Directors of Education Pedagogical Commission, whose aim is to coordinate common standards, textbooks, and manuals across the country; teacher salaries are very high, and the occupation has much respect; and university experts supervise the examination process.

Contrasting Flemish Belgium and the Netherlands. Flemish Belgium is unique in our focus group of countries in that it does not have high-stakes exit examinations. To maintain coherence, they must control quality at the front end of the process. By contrast, the character of the education system in the Netherlands requires that quality control be maintained at the back end of the curriculum and instruction process.

In Flemish Belgium, the expert respondent claimed a "100% match" between the content of textbooks and teaching materials and the content standards. The textbooks are written by the same people who develop the curriculum guides. The curriculum objectives are made public by the media and through public relations campaigns of the education ministry, complete with leaflets and brochures printed on a large scale and disseminated widely. With this, parents and the public can better judge their schools' performance because they can know what they are supposed to be teaching. Curriculum-based inspections are pervasive and are used to see if teachers are teaching the correct material and doing it on time, although it has been proposed that inspections be done only at the school level rather than at the classroom level.

That Flemish Belgium does not have high-stakes exit examinations does not, in itself, mean that students never risk rejection. Flemish Belgium maintains separate upper secondary level curricular tracks, some of which are highly selective. Getting into the track of one's choice may require a better school record than those of other students who wish entry into the same track. Moreover, teachers can still fail students, even without high-stakes standardized tests, and indeed, some educators in Flemish Belgium perceive a problem of too many grade repeaters at the upper secondary level.

In the neighboring Netherlands, one could describe the structure of the quality control system as the converse of Flemish Belgium's. The Netherlands maintains a very open system of school choice and a great variety of schools. There are Catholic schools, Protestant schools, Islamic schools, and "Green" schools; virtually any group can start a school and receive full public funding. These schools use a wide variety of textbooks and curriculum materials. Schools can choose their own curriculum, and the implementation of curricula is unsupervised by the government. Indeed, the national constitution prevents the establishment of an official curriculum.

The national government does offer guidance on a voluntary basis, maintaining local and regional advisory guidance centers, a national Curriculum Development Institute, a semiautonomous test development organization, tight subject-area networks of teachers who help to develop and score examinations, and university departments that have taken over some quality control functions within each respective subject area.

Also, the Netherlands administers high-stakes standardized examinations, prominently and frequently. The government allows much public input as to the content of the examinations, and topics that are culturally sensitive (e.g., evolution) might not be included. But once the content domain of the examinations is set, schools are required to administer them, and students are required to pass them.

As one spokesperson has written (*Encyclopedia of Comparative* 1988, 504),

> The strongly differentiated Dutch system requires a radical decision about every pupil at the end of every school phase, a decision which, to a large extent, determines the pupil's future profession, income, and social standing.
>
> Promotion from grade to grade in primary school is decided by norm-referenced tests. Those in the bottom quartile are not promoted while the others are.

Like the education system in Flemish Belgium, the Netherlands' also creates a high number of failing students, which worries some educators.

COMPARING QUALITY CONTROL IN HIGH-ACHIEVING COUNTRIES TO THAT IN THE UNITED STATES

There are some characteristics of the curriculum and instruction quality control systems common to all or most of the countries in our focus group that contrast markedly with systems common in the United States.[2]

1. CLASSROOM- AND CURRICULUM-BASED INSPECTIONS

In the United States, school inspections are infrequent and are done on a schoolwide basis, with the school as a whole attaining or not attaining accreditation based on schoolwide measures of inputs or performance. In some of our group of high-achieving countries, classroom-level and/or curriculum-based inspections also exist.

It is more common in our focus group of high-achieving countries to find the systemwide responsibility for curriculum and instruction quality control assumed by subject-area experts. In mathematics, this usually means mathematics professors at universities or mathematicians in the education ministries. This stands in contrast to the typical situation in the United States where there are few mathematics experts in state education agencies or local school districts, and they are likely education school rather than mathematics department graduates. Most university mathematics departments in the United States have no connection or involvement in mathematics teaching at the primary and secondary levels.

2. CONTENT STANDARDS THAT ARE FIXED AND EXPECTED TO BE FOLLOWED AS A CORE CURRICULUM

These curriculum-based inspections in our focus group of high-achieving countries can be rather standardized because, everywhere but the Netherlands, teachers are expected to follow a common curriculum according to a common timetable. The inspectors, then, can judge the teacher against a common curricular standard. In the United States, curricula and texts are so diverse and timetables so anomalous that it would be difficult to conduct a classroom-level, curriculum-based inspection. How would the teacher's performance be measured? There is no clear standard.

What happens to teachers in these high-achieving countries who deviate from the standard program? One of our respondents asserted, "They do not deviate." The common curriculum typically occupies 80% to 100% of the

instructional time. Our respondents in Singapore, France, and the Czech Republic pointed out that teachers were free to depart from the common curriculum if their class was ahead of schedule; they wished to provide practical, everyday examples of abstract content; or they wanted to use examples from magazines or videos to motivate interest. But in all countries, students would still be held accountable for mastering the core curriculum.

3. MORE HIGH-STAKES SELECTION POINTS

Most of our high-achieving countries have few, several, or many high-stakes selection points. Most administer one, two, three, or several high-stakes entrance or exit examinations. Most are also selective in their admissions to certain programs or curricular tracks, with low-achieving students at one level of education denied their first choice of curricular track at the next level of education. Flemish Belgium is unique in lacking the examinations, but they still maintain selective admissions to certain programs and curricular tracks, selective based on academic performance.

This stands in contrast to the United States, where most states with high-stakes examinations have only low-level "minimum competency" literacy tests for high school graduation. Curricular tracking is also uncommon. Only in the small proportion of school districts with magnet programs or career academies with selective admissions do such stakes apply in the United States.

4. EXAMINATIONS THAT ARE CURRICULUM-BASED AND HIGH STAKES

U.S. states with low-level "minimum competency" literacy tests for high school graduation may be said to have high-stakes curriculum-based tests, but they are genuinely of high stakes only for a small proportion of students at risk of failing them, and they are typically based on curriculum from the primary or lower secondary level. Take away minimum competency tests and few U.S. states have high-stakes tests. A study by the U.S. General Accounting Office in 1993 concluded that only one quarter of tests administered district- wide in the United States had high stakes for students. The large majority of them were statewide minimum competency tests. Surely, that proportion is higher now but still not as high as in most European countries.

High-achieving countries tend to have high-stakes examinations of some variety—at varying levels of difficulty or in different curricular tracks. Singapore offers the British-inspired "O" level (*O* is for "ordinary") and "A" level (*A* is for "advanced") examinations. France requires passage of exit

examinations in several academic tracks of differing curricular emphases (e.g., language and humanities, natural science, physical science and mathematics, economics, technology), as well as some vocational and professional tracks.

5. SECONDARY SCHOOLS ORGANIZED BY CURRICULAR FOCUS

Organizing secondary schools by curricular focus can aid quality control because it helps to focus the efforts of those authorities responsible for monitoring curriculum. A French inspector, expert in the math/physics/chemistry curriculum series can attend classes in that subset of schools that offer this curriculum series. Curriculum experts at the national ministry, likewise, can specialize in that particular mathematics curriculum and focus on those particular schools.

6. OTHER PRACTICES THAT REINFORCE COHERENCE

Other practices that reinforce coherence and are common in our group of high-performing countries but not in the United States include the following: high school-level standards for promotion to the next grade, as evidenced by a relatively high rate of *redoublement*, or retention in grade; ability grouping; passage of subject-area standardized tests required of teachers; "looping" (i.e., teachers in lower grades may keep the same group of students for multiple years and thus are held more accountable and have an incentive to make certain all students make reasonable progress); and employers' use of grades or test scores in their hiring decisions, reinforcing the importance of studying.

DECISION POINTS

A country may profess to many methods of quality control, but if there are no consequences for a failure to adhere to them, they may well be ignored.

Thus, another way to contrast different countries' quality control systems for curriculum and instruction is to identify the type and number of decision points, or quality control measures, where adherence to the curriculum and instruction system can be reinforced. Most decision points involve stakes for the student, teacher, or school. They involve potential consequences for failure to adhere to the system and to follow the program at a reasonable pace. Students may be denied promotion if they do not study. Teachers may be

denied employment if they do not pass exams demonstrating subject-area expertise. Schools may suffer sanctions if it is shown that their students are not keeping up with their studies or studying the correct materials.

DECISION POINTS OF TOP-PERFORMING COUNTRIES

Table 3 contrasts the decision points used in the focus group of countries to those used in the United States. "Yes" is written if a country used a certain decision point to monitor or maintain coherence to a curriculum and instruction system, "No" is written if it could be determined that a country did not use that decision point, and blank cells indicate no information was found for that country during the study.[3] Most decision points involve selection; some students or teachers are or are not selected if they do or do not maintain adherence to the program.

Table 3 consists only of "systemwide" decision points—those universally maintained. Nonsystemwide or local decision points are those that are enforced only at the local, school, or classroom level, such as retention in grade.

Counting the number of "Yes" cells that indicate the existence of a decision point, one can see that each of the focus group countries maintains 10 or more decision points, while the United States maintains 6. The category "Some" was counted as one half. The mean number of decision points among the top-performing countries is 13.88, more than double the United States' 6.

Comparing the average number of systemwide decision points of the top-performing countries (13.88) to the United States' 6, one finds the U.S. total to be more than 2 standard deviations ($s = 3.14$) below the top-performers' average.

Table 4 contrasts the prevalence of the local decision points of retention in grade among the focus group of countries and the United States. The average rates of retention in grade for the focus group of countries were 0.86 students per school for Grade 4 and 2.54 students per school for Grade 8 (rates are listed for each country and each grade level in "Note" under Table 4). For the United States, the rate of retention was higher for Grade 4 (1.01) and lower for Grade 8 (1.65) (TIMSS, unpublished computations). More than half the total average number of students retained for the eight countries comes from France.

The U.S. rate of retention in grade was not significantly different than the top-performing countries' rates. (Some readers may be tempted to assume from looking at Table 4 that low retention rates are the norm for East Asia; rates range from 0 to 0.6 in Japan, Korea, and Singapore. To provide some

(Text continues on p. 414)

TABLE 3: Systemwide Decision Points (activities with stakes and consequences for student, teacher, or school), by Country: 1994-1995

	Belgium, Flemish	*Czech Republic*	*France*	*Korea*	*Japan*	*The Netherlands*	*Singapore*	*Switzerland*	*United States*
Level of education exit exam									
Primary level	No	No	No	No	No	Yes[8]	Yes[2,3,6,7]	Yes[8,21]	No
Lower secondary	No	No	Yes[3,4]	No	Yes[21]	Yes[3]	Yes[3,7,a]	Yes[8,21]	No
Upper secondary	No	Yes[1,2,21]	Yes[1,4]	Yes[1]	Yes[1,a]	Yes[1,3,5]	Yes[1,3,6,7]	Yes[1,8,21]	Some
Level of education entrance exam									
Lower secondary	No	Yes[9]	No	No	Yes[21]		Yes[6]	Yes[11,21]	No
Upper secondary	No	Yes[2,3,7,9,21]	Yes[8]	Yes[3,10]	Yes[4,8,10,a]	Yes[21]	Yes[6]	Yes[8,21]	No
Higher education	Yes[7]	Yes[2,5,7,9,21]	Yes[3,4]	Yes[3,21]	Yes[3,4,a]	Yes[21]	Yes[3,6]	Yes[8,21]	Yes
Other types of standardized exams									
Assessments	Yes[12]	No	Yes[2,7,11]	Yes[3,21]	Yes[12]	Yes[10]	Yes[15]	Yes[8]	Yes
End-of-course	No	No	No	Yes[21]		Yes[10]	Yes[15]	Yes[21]	No
Others				Yes[21]	Yes[21]	Yes[21]	Yes[21]	Yes[21]	Yes
Selection of schools or students for certain curricular tracks									
Lower secondary	Yes[21,a]	Yes[9,11]	Yes[10]	No	No	Yes[10,15]	Yes[6,12,a]	Yes[10,21]	No
Upper secondary	Yes[7,21]	Yes[10]	Yes[3,10,15]	Yes[3]	Yes[3,a]	Yes[10,15]	Yes[12,a]	Yes[10,15,21,a]	No
Higher education	Yes[7]	Yes[9,11]	Yes[10]	Yes[3]	Yes[3,a]	Yes[21]	Yes[6,12,a]	Yes[10,15,21,a]	Yes
Ability grouping common within schools									
Primary level[21]	No			No	No		Yes		Some
Lower secondary[21]	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Some
Upper secondary[21]	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes

(continued)

TABLE 3 Continued

	Belgium, Flemish	*Czech Republic*	*France*	*Korea*	*Japan*	*The Netherlands*	*Singapore*	*Switzerland*	*United States*
Large nonpublic sector makes more school selection possible (> 25%)									
Primary level	Yes[21]	No[21]	No[21]	No[21]	No[21]	Yes[1,21]	Yes[1,21]	No[21]	No
Secondary level	Yes[1]	No[9]	Yes[1,21]	Yes[1,21]	Yes[1,21]	Yes[1,21]	Yes[1,21]	Yes[1,21]	No
School system and classroom practices									
Classroom instruction is inspected	Yes[a]	Yes[9]	Yes[14]	Yes[3]	Yes[a]		Yes[a]	Yes	No
Examination required in subject area for teachers[1]	Yes	Yes	Yes	Yes[1,3]	Yes	Yes	Yes	Yes	No[1]
Total "Yes"[b]	11	10	13	12	13	16	19	17	6

NOTE: 1. Beaton (1996); 2. Bishop (1997); 3. Postlethwaite (1996); 4. Stevenson and Lee (1997); 5. Peak (1997); 6. Yeoh (1996); 7. Postlethwaite (1988); 8. Phelps (1996); 9. Organisation for Economic Co-operation and Development, "Czech Republic"; 10. Schmidt; 11. Kreeft (1990); 12. Phelps (2000); 13. National Center on Education and the Economy (1994); 14. Resnick, Nolan, and Resnick (1995); 15. Bishop (1993); 16. Organisation for Economic Co-operation and Development, "France"; 18. U.S. Department of Education (1992); 19. Organisation for Economic Co-operation and Development, "Belgium"; 20. Resnick, Nolan, and Resnick (1995); 21. Robitaille (1997); 22. Third International Mathematics and Science Study, unpublished computations; 23. Organisation for Economic Co-operation and Development, "Spain"; 24. Asia-Pacific Economic Cooperation (1998); 25. Organisation for Economic Co-operation and Development, *Investing in Education: Analysis of the 1999 World Education Indicators* (2000); 26. Organisation for Economic Co-operation and Development, "Greece."

a. Source is response to this study's survey.

b. Scoring: yes = 1, some = 0.5, no = 0.

TABLE 4: Local Decision Points (activities with stakes and consequences for student, teacher, or school), by Country: 1994-1995

	Belgium, Flemish	*Czech Republic*	*France*	*Korea*	*Japan*	*The Netherlands*	*Singapore*	*Switzerland*	*United States*
Retention in grade is common									
Primary level[22] (> 3%)	Yes[19]	No	Yes[3,15,22]	No	No	Yes	No	Yes[3,22]	No
Secondary level[22](> 5%)	No[19,22]	No	Yes[3,15,22]	No	No	Yes	No	No	No
Total "Yes"[a]	1		2			2		1	

NOTE: On citations and superscripts, blank cell means no information found or not applicable; cell (and row) with no superscript means no information source declares the information, but a lack of information to the contrary from several sources implies it, or the information is common knowledge; superscript for row title means all cells have information from the same source document, unless otherwise indicated in the cell. Mean rate for 3rd and 4th grades (Czech Republic, 1.00; Japan, 0.0; Korea, 0.14; the Netherlands, 3.16; Singapore, 0.02) = 0.86 students per grade per school; U.S. rate = 1.01 students; mean rate for 7th and 8th grades (Belgium, 2.95; Czech Republic, 1.19; France, 10.33; Japan, 0.0; Korea, 0.06; the Netherlands, 3.29; Singapore, 0.6; Switzerland, 1.93) = 2.54 students per grade per school; U.S. rate = 1.65 students; overall mean rate = 1.70 students, U.S. rate = 1.33 students. 3. Postlethwaite (1996); 10. Schmidt; 15. Bishop (1993); 22. Third International Mathematics and Science Study, unpublished computations.

a. Scoring: yes = 1, no = 0.

perspective, however, Hong Kong retains 1.58 students per school per year in 4th grade and 2.71 students per school per year in 8th grade.)

Thus far, we have seen that the United States is different. It seems to maintain less quality control over its curriculum and instruction system than do the top performers in the TIMSS. For all we know, however, the United States may be different from most other countries, regardless of whether they are top performers. If the bottom performers in the TIMSS also use more quality control measures than the United States, we will have learned nothing about the relationship between quality control and student achievement.

To check this possibility, information adequate to fill in tables like the two immediately above was gathered for the bottom performers in the TIMSS.

DECISION POINTS OF THE BOTTOM PERFORMERS IN THE TIMSS

Again, in Table 5, we contrast a focus group of sorts, the dozen countries scoring worst on the TIMSS. In this case, we get quite different results. The total number of quality control measures ranges from two to seven. The countries with the most quality control measures in this list, Iran and Latvia, still use three fewer than the country in the "top performers" focus group with the fewest measures. The United States, with six quality control measures, fits right into this group of bottom performers, tied with Germany and the Philippines.

Comparing the average number of systemwide decision points of the bottom-performing countries (4.42) to the United States' 6, one finds the U.S. total to be between 1 and 2 standard deviations ($s = 1.88$) above the bottom-performers' average. The average number of decision points of the bottom-performing group is statistically significantly different from that of the top-performing group, as determined by a t test ($t = 7.69$, $p < .0001$) between the two means of 13.88 and 4.42 ($s = 3.14$).

(Some readers may be tempted to assume from looking at Table 5 that Mediterranean countries tend to use few quality control measures; Cyprus, Greece, Portugal, and Spain represent four of the five countries with the fewest measures used. To provide some perspective, however, Italy, which did not participate in the TIMSS, is a Mediterranean country that requires passage of high-stakes examinations at three different levels of education and selection to curricular tracks at both secondary levels. Italy offers a rigorous system with a relatively high number of decision points; thus, the Mediterranean climate does not necessitate a lack of rigor.)

The average rates of retention in grade for the focus group of countries (see Table 6) were 3.89 students per school for Grade 4 and 6.34 students

(Text continues on p. 418)

TABLE 5: Systemwide Decision Points (activities with stakes and consequences for student, teacher, or school), by Country: 1994-1995

	Columbia	*Cyprus*	*Germany*	*Greece*	*Iceland*	*Iran*	*Latvia*	*Lithuania*	*Philippines*	*Portugal*	*Romania*	*Spain*	*United States*
Level-of-education exit exam													
Primary level	No[3]	No[3]	No	No[26]	No[3]	No[3]	No[3]	No[3]	No[24]	No[3]	No[3]	No	No
Lower secondary	No[3]	No[21]	No	No[26]	No[3,21]	Yes[3,21]	Yes[3,21]	Yes[3,21]	Some[24]	Yes[3]	No[3]	Yes[3]	No
Upper secondary	No[2]	No[2,3]	Yes	No[3]	No[3]	Yes[2,3]	Yes[3,21]	Yes[3,21]	Some[2,21]	Yes[3]	No[3]	No[3]	Some
Level-of-education entrance exam													
Lower secondary	No[3]	No[3]	No	No[26]	No[3]	No[3]	No[3]	No[3]	No[24]	No[3]	No[3]	No[3]	No
Upper secondary	Some[21]	No[3]	No	No[26]	No[3]	Yes[3]	Yes[3]	No[3]	No[24]	No[3]	Yes[3]	No[3]	No
Higher education	Some[21]	No[3]	Yes	Yes	No[3]	Yes[3]	No[3]	No[3]	Some[21,24]	Yes[3]	Yes[3]	Yes[3]	Yes
Other types of standardized exams													
Assessments	No	No[3]	No	No	No	No[3]	Yes[21]	No[3,21]	Yes[21]	No	No	No	Yes
End-of-course	No	No[3]	No	No	No	No[3]	No[3]	No[3]	Yes[21]	No	No	No	No
Others	No	No	No	No	No	No	No	No	No	No	No	No	Yes
Selection of schools or students for certain curricular tracks													
Lower secondary	No[21]	No[3]	Yes	No[3]	No	No[3]	No[3]	No[3]	No[24]	No	No	No	No
Upper secondary	No[21]	No[3]		No[3]	No	Yes[3,21]	Yes[3]	No[21]	No[25]			Yes	No
Higher education	Yes[3]	No[3]	Yes	Yes[3]	No	Yes[3]	Yes[3]	Yes[3]	Yes[24]				Some
Ability grouping common within schools													
Primary level[21]	No	No	No	No	No	No	No	No	No			No	Some
Lower secondary[21]	No	No		No	Some	No		Some	No		Yes		Some
Upper secondary[21]	Yes	Yes		No	Some	Yes		Yes			Yes		Yes
Large nonpublic sector makes more school selection possible (> 25%)													
Primary level[21]	No	No	No	No	No	No	No	No	No		No		No
Secondary level[21]	Yes[3]	No[3,22]	No	No[3,22]	No	No	No[3,22]	No[3,22]	Yes		No	No[23]	No

(continued)

TABLE 5 Continued

	Columbia	*Cyprus*	*Germany*	*Greece*	*Iceland*	*Iran*	*Latvia*	*Lithuania*	*Philippines*	*Portugal*	*Romania*	*Spain*	*United States*
School system and classroom practices													
Classroom instruction inspected[3]	No[3,21]	Yes	Yes	No	No		No	Yes			Some		No
Teacher exam in subject area required[1]	No	No	Yes	No	Yes	No	Yes	No[1,3]	Yes	No	Yes	Yes	Some
Total "Yes"[a]	4	2	6	2	2	7	7	5	6	3	5	4	6

NOTE: On citations and superscripts, blank cell means no information found or not applicable; cell (and row) with no superscript means no information source declares the information, but a lack of information to the contrary from several sources implies it, or the information is common knowledge; superscript for row title means all cells have information from the same source document, unless otherwise indicated in the cell. 1. Beaton (1996); 2. Bishop (1997); 3. Postlethwaite (1996); 4. Stevenson and Lee (1997); 5. Peak (1997); 6. Yeoh (1996); 7. Postlethwaite (1988); 8. Phelps (1996); 9. Organisation for Economic Co-operation and Development, "Czech Republic"; 10. Schmidt; 11. Kreeft (1990); 12. Phelps (2000); 13. National Center on Education and the Economy; 14. Resnick, Nolan, and Resnick (1995); 15. Bishop (1993); 16. Organisation for Economic Co-operation and Development, "France"; 18. U.S. Department of Education (1992); 19. Organisation for Economic Co-operation and Development, "Belgium"; 20. Resnick, Nolan, and Resnick (1995); 21. Robitaille (1997); 22. Third International Mathematics and Science Study, unpublished computations; 23. Organisation for Economic Co-operation and Development, "Spain"; 24. Asia-Pacific Economic Cooperation (1998); 25. Organisation for Economic Co-operation and Development, *Investing in Education: Analysis of the 1999 World Education Indicators* (2000); 26. Organisation for Economic Co-operation and Development, "Greece."
a. Scoring: yes = 1, some = 0.5, no = 0.

TABLE 6: Local Decision Points (activities with stakes and consequences for student, teacher, or school), by Country: 1994-1995

	Columbia	*Cyprus*	*Germany*	*Greece*	*Iceland*	*Iran*	*Latvia*	*Lithuania*	*Philippines*	*Portugal*	*Romania*	*Spain*	*United States*
Retention in grade is common													
Primary (> 3%)[22]	Yes[3,22]	No[3,22]	No[5]	No	No	Yes	Yes			Yes			No
Lower secondary (> 5%)[22]	Yes[3,22]	Yes[22]	Yes	Yes	No	Yes	Yes	No		Yes	No	Yes	No
Total "Yes"[a]	2	1	1	1		2	2			2		1	

NOTE: On citations and superscripts, blank cell means no information found or not applicable; cell (and row) with no superscript means no information source declares the information, but a lack of information to the contrary from several sources implies it, or the information is common knowledge; superscript for row title means all cells have information from the same source document, unless otherwise indicated in the cell. Not all focus group countries filled in this information in the school background questionnaires (i.e., the Philippines did not); mean rate for 3rd and 4th grades (Cyprus, 1.0; Greece, 1.0; Iceland, 0.54; Iran, 4.92; Latvia, 4.54; Portugal, 10.41) = 3.89 students per grade per school; U.S. rate = 1.01 students; mean rate for 7th and 8th grades (Columbia, 8.46; Cyprus, 4.56; Germany, 6.06; Greece, 8.81; Iceland, 1.55; Iran, 11.66; Latvia, 3.69; Lithuania, 2.82; Portugal, 8.39; Romania, 2.94; Spain, 10.78) = 6.34 students; U.S. rate = 1.65 students; overall mean rate = 5.12 students per grade per school; U.S. rate = 1.33 students. 3. Postlethwaite (1996); 5. Peak (1997); 22. Third International Mathematics and Science Study, unpublished computations.

a. Scoring: yes = 1, some = 0.5, no = 0.

TABLE 7: Summary of Decision Point Information

	Top Performers (mean)	*Bottom Performers (mean)*	*United States*
Systemwide measures			
Number of decision points	13.88	4.42	6
Local measure			
Number of decision points	0.75	1.00	0
Rate of retention in grade (percentage) (Grades 7 and 8)	2.34	6.34	1.65

per school for Grade 8 (rates are listed for each country and each grade level in "Note" under Table 6). For the United States, the rates of retention were lower for Grade 4 (1.01) and for Grade 8 (1.65) (TIMSS, unpublished computations).

Comparing the average rate of retention in the 7th and 8th grades among the bottom-performing countries (6.34) to the United States' 1.65, one finds the U.S. rate to be between 1 and 2 standard deviations ($s = 3.46$) below the bottom-performers' average. The average rate of retention of the bottom-performing group is significantly different from that of the top-performing group, however, as determined by a two-tailed t test ($t = 2.39$, $p < .05$) between the two means of 2.54 and 6.34 ($s = 3.38$).

DECISION POINTS: SUMMARY

Table 7 displays a concise summary of the decision point discussion. The United States uses fewer quality control measures (i.e., decision points) systemwide than top-performing countries do, but slightly more than bottom-performing countries use, on average. The United States, on average, has a low rate of retention in grade (1.65 students per class per year for 7th and 8th grades and 1.33 for both primary and secondary school), the single example of local quality control measure used in this analysis. Top-performing countries have a somewhat higher rate of retention in grade, whereas bottom-performing countries have a much higher average rate of retention in grade (6.34 students per class per year in Grades 7 and 8, and 5.12 for both primary and secondary school).

Figure 1 contrasts the top- and bottom-performing groups of countries (here, the United States fits neatly into the bottom group) on the relationship between their number of systemwide decision points and average percentage of correct answers on the 7th and 8th grade level TIMSS tests. The scatterplot

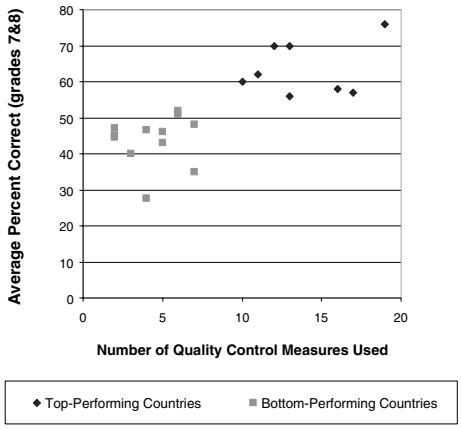

Figure 1: Average Third International Mathematics and Science Study Score and Number of Quality Control Measures Used, by Country

Country	*Decision Points*	*Seventh and Eighth Grade Average Percentage Correct*
Singapore	19	76
Switzerland	17	57
The Netherlands	16	58
Japan	13	70
France	13	56
Korea	12	70
Czech Republic	11	62
Belgium	10	60
Latvia	7	48
Iran	7	35
Germany	6	52
United States	6	51
Lithuania	5	43
Romania	5	46
Spain	4	46.5
Columbia	4	27.5
Portugal	3	40
Iceland	2	47
Greece	2	44.5
Cyprus	2	45

$p = .776712$.

implies a positive relationship between more quality control measures enforced (i.e., decision points) and higher test scores (the Pearson product-moment correlation is 0.78).

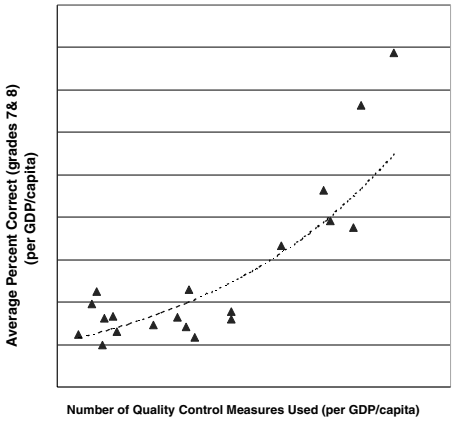

Figure 2: Average TIMSS Score and Number of Quality Control Measures Used (each adjusted for GDP/capita), by Country
NOTE: TIMSS = Third International Mathematics and Science Study.

A skeptic might speculate that wealthier countries have a considerable advantage in promoting student achievement, such that country wealth might be the key driver of achievement, not quality control measures, or anything else. Indeed, there does appear to be some correlation ($p = .54$) between countries' 8th-grade TIMSS mathematics scores and their GDP per capita. More to the point, however, if the implementation of quality control procedures requires more resources, and quality control procedures improve student achievement, then is it not really wealth that is improving student achievement? The Pearson product-moment correlation coefficient between the number of quality control measures (i.e., decision points) used and GDP per capita is 0.47 for the group of countries included here.

In order to adjust for country wealth, then, both of the factors deployed in Figure 1 were divided by GDP per capita. The derived factors are measures of test scores and quality control procedures per unit of wealth (i.e., average percent correct [TIMSS 8th-grade math] per GDP per capita, and number of quality control measures used per GDP per capita). With the factor of wealth removed, do we still find a positive correlation between student achievement and quality control? Indeed, we do; see Figure 2.

Figure 2 suggests an exponential relationship between quality control measures and student achievement. It would appear that, up to a certain point, quality control implementation makes some difference in student achievement, even when the resources available for quality control implementation are taken into account. But, after that point, if an extra effort is made to imple-

ment quality control procedures in spite of limited resources, student achievement can really take off.

Judging from all the information considered thus far related to the prevalence of decision points (a.k.a., quality control measures), it would appear,

- Top-performing countries use more systemwide quality control measures. The U.S. number lies in between the averages of the top and bottom performers but is closer to the bottom.
- The bottom performers use more of the local quality control measure, retention in grade, perhaps as a substitute for the systemwide measures they lack.
- The United States is low on all summary statistics—closer to the bottom performers on systemwide measures and lower than both top- and bottom-performing countries on local measures.

Opponents of local quality control measures, such as retention in grade, perhaps, could increase their chances of achieving its abolition if they advocated for more systemwide measures of quality control, such as high-stakes tests. It would appear that the presence of an integrated system of systemwide quality control measures might reduce the need for local control measures.

COMPARISONS TO THE U.S. SYSTEM

Although one can observe a good deal of similarity in curriculum among U.S. classrooms, there is little uniformity. U.S. textbooks in 1994-1995, for example, share a large degree of similarity in appearance and content but are not deliberately alike and not alike enough to represent a common curriculum or to form a common item pool for high-stakes testing at more than a minimal level of competency. Some even argue that they are "dumbed down" to a lowest common denominator to be salable to the largest possible population of classrooms. Moreover, there is no assurance in most of the United States, even with common textbooks, that two teachers in different classrooms are interpreting the content the same way, at the same pace, or even at the same grade level.

One might argue that the United States benefits from a great diversity in curriculum and instruction. One defense of the U.S. system might be that if different students learn different content, then the country as a whole benefits because no matter what the topic, we are more likely to have citizens who possess the knowledge, than are other countries where all their citizens learn all the same content. Another defense is that each teacher gets to tailor

curriculum and instruction to his or her own particular strengths and to his or her students' particular needs.

Critical responses to the first defense could include the following: The U.S. curriculum actually appears to be burdened with a great deal of repetition and superficiality (see McKnight et al. 1987; Schmidt et al. 1996a). Another response is that there is a great deal of variety in curriculum and instruction in high-performing countries, too, but it is organized more rationally. Separate schools exist with curricular focus and students who wish to share that focus attempt to enter those schools.

The second defense of the U.S. system—about tailoring classroom curriculum and instruction to the personal characteristics of the teacher and the students—is heard often. Most of our expert respondents from top-performing countries thought that it was important that teachers have some flexibility to tailor curriculum and instruction to their classes. To do that, the required core curriculum typically takes up only 80% of classroom time. A buffer of 20% of the school year is conserved to allow slower moving classes to catch up with their faster moving colleagues by the end of each school session. The faster moving classes use the buffer time for enrichment exercises, such as exercises in the practical applications of mathematics concepts in real life, with examples provided from daily life or the popular press. So these top-performing countries typically do not demand 100% uniformity, only 80%. One could argue that in the United States, the equivalent figure is 0%.[4]

The most commonly experienced drawback to the heterogeneity of curriculum in the United States is experienced by the children of families who move. These children can discover that in their new school district, they are behind schedule, ahead of schedule, not prepared, overly prepared, and so on. Commonly, they enter a completely different curriculum in the absence of common system standards, and they waste time. For kids in families that move often, the kids can suffer academically. In France, with its uniform curriculum nationwide, there simply is no such problem.

This examination of quality control over curriculum and instruction in top-performing countries suggests another drawback. Without common, enforceable standards, there may be no good way to affect performance systemwide other than through high-stakes standardized tests (as in the Netherlands). Without either common standards or high-stakes standardized tests, there may be no effective way at all to monitor performance systemwide. Some U.S. teachers may be doing a wonderful job in their totally customized classes, but some may be doing an awful job. How is one to know or tell which?

In the United States, one must hope that teachers will face down the natural incentives of their students, parents, schools, and themselves to avoid

accountability by holding themselves and their students to high standards of performance. One must also hope that teachers will know how.

The tight networks of subject-area professionals in top-performing countries provide classroom-level inspections. Some teachers might feel threatened by these inspections, but they might also benefit from advice the inspectors have to offer. With a common core curriculum, inspectors can offer advice from a deep pool of knowledge about what works, because all teachers are teaching the same material. With no common core curriculum in the United States and every class taught in a unique, customized manner, any classroom-level curriculum-expert inspectors, were there to be any in the United States, would have less to say, and it would be less specific.

In one nationwide survey of U.S. teachers, 99% responded that they thought subject matter knowledge *should be* considered in their performance evaluation, whereas only 65% said it *was* (Nolan 1997, iii, 8, 27). Even then, where performance evaluations are conducted by school principals, odds are that the principal is not expert in most teachers' subject matter.

LESSONS FOR THE UNITED STATES

Top-performing countries tend to use a lot of quality control measures, such as high-stakes examinations, selection for curricular tracks, ability grouping, and other devices considered anathema by many U.S. education professors. The "progressives" in the United States who oppose testing, tracking, and ability grouping may wish to ignore most of the top-performing countries and embrace Flemish Belgium for solace.

How much will they find? If they are honest with themselves, not much. First, the Flemish community of Belgium uses ability grouping and selection for curricular tracks; it is only high-stakes tests that they do not use or, rather, did not use until the late 1990s, when they started development of an upper secondary school exit exam.

Second, Flemish Belgium is just one country, alone among the top-performing countries in its absence of high-stakes examinations. Most countries eschewing high-stakes tests scored poorly on the TIMSS.

Third, Flemish Belgium does not compare well to the larger U.S. states; it is just too small. Some of its key quality control features, such as the constant and close interaction of teachers, and the highly visible public dissemination of information on standards, are probably easier to implement in smaller jurisdictions.

Nonetheless, progressives may wish to look to some U.S. adaptations of the Flemish Belgium sort. For a state model, they could look to Connecticut, which attempts to maximize the amount and the public visibility of information on school and student performance without using high-stakes examinations. They do, however, contrary to the Flemish Belgium of 1995, administer lots of standardized tests, but although some of those tests have stakes for the students—e.g., the 4th, 6th, and 8th grade Mastery Tests—Connecticut has no high-stakes exit examination.

Moreover, Connecticut maintains some other quality control features similar to those found among the TIMSS top performers:

- Connecticut is one of the few U.S. states to have long retained a detailed state curriculum, in place long before the current standards movements, that was taken seriously by local school districts.
- Connecticut employs "master" teachers to review and critique new teachers in the classroom. New teachers are reviewed often, through direct classroom observation and videotape. Critical evaluations from master teachers can cost new teachers their jobs.
- The state publishes a statewide report card that compares districts on a number of indicators of progress and success (or, lack thereof).

For a model quality control measure, progressives may wish to look to the use of school and district report cards in the United States. Statistical correlations between improvement over time on state National Assessment of Educational Progress (NAEP) scores and the existence of school and district report cards in the state are as strong as the correlations between the existence of state high-stakes tests and improvement on state NAEP scores. This suggests that public glory and embarrassment may be as effective a quality control inducement as the genuine consequences of high-stakes testing.

I suspect, however, that many U.S. progressives would not accede even to the use of school and district report cards or high-stakes master teacher evaluations; such behavior runs counter to the beliefs of more radical constructivists and egalitarians, who would regard both as invalid and unfair.

What are the lessons of this study for those progressives who want no decision points and no quality control measures? Be prepared to accept last place in the Fourth International Mathematics and Science Study, below Cyprus and Greece (Iceland is currently busy building a rather comprehensive examination system, from scratch). If their goals for the U.S. education system tend toward what they regard to be noble public goods, such as the impartation of beliefs in egalitarianism and their version of moral and civic consciousness, they may genuinely not care that U.S. academic achievement dives toward the bottom.

What is the lesson of this study for the "traditionalists?" Probably, it confirms what they have suspected all along. How much of a thorough, integrated quality control system do we see in the United States? Are we at least heading in the direction of building such a system? For at least half of the measures, yes.

Many U.S. states are now in the process of implementing systemwide quality control measures (i.e., decision points). More than a few states have or are implementing high-stakes examinations at several levels. Some states have or are implementing examinations at the same level with more than one level of difficulty, for a regular diploma and an honors diploma (e.g., New York). Some states have or are implementing curricular choices in those exit exams (e.g., passage of 5 subject-area exams among a choice of 10), and those choices may eventually lead to the adoption of curricular tracking. Ability grouping is already common in most of the United States, although many education professors claim that "the research" shows it to be a bad thing.

Subject-area mastery for teachers, with education-school exit exams based on subject-area knowledge as well as pedagogical concepts, is fast becoming a standard requirement in the United States.

The remaining aspects of fully integrated quality control systems may still elude U.S. school systems for some time to come. We may never see classroom- and curriculum-based teacher instruction inspections to the degree that they exist in other countries. Such systems would need to be built from scratch. Some states have been experimenting with programs that promote the best to be "master" teachers, who no longer teach a full class load themselves but visit other teachers' classrooms and give them advice. But few states are as far along in using this technique as Connecticut, which uses it only with new teachers.

More likely, it would appear based on current trends, that teachers will be judged based on their students' gains, in scores on curriculum-based tests. In the examination systems most fair to teachers (e.g., Tennessee), student test scores are adjusted for background factors, such as demographic profiles, and the students are tested frequently, so that the pressure is distributed across teachers in all grades, not just a few testing grades.

Given the choice, teachers would probably prefer classroom-based inspections. Indeed, when former president Al Shanker was urging his American Federation of Teachers to enthusiastically support high standards and high-stakes standardized tests, he often cited European countries as a model. There, he found high standards, high-stakes tests (for students) and high levels of professionalism in classroom instruction and school administration, alongside teacher corps that were completely unionized, highly paid, and

high in social status (the latter point quite a contrast to most of the United States).

Separating classrooms and schools along the lines of different curricular tracks may be difficult to implement in the United States and encounter much opposition. It might seem antidemocratic to some. If the charter school movement really takes hold, however, the adoption of curricular tracking within and by schools will only be a matter of time. If parents and students are given a choice, most will probably choose some clear curricular or occupational direction over the current bland generality. Even in the public school systems, career academies and magnet schools already offer curricular tracking, and many of these programs are very selective.

CONCLUSION

All other factors being equal, quality control must be more difficult in the absence of common standards. This study of top-performing countries suggests that the most successful quality control efforts manage rather thoroughly the entire chain of elements that make up the curriculum and instruction system.

An interesting study managed by David Cohen at Michigan State University tells the story of a Michigan State effort to change curriculum and instruction in mathematics through a standardized program. Very careful and thorough, the program seemed to consider every essential aspect. The story follows activities at the state level, public relations level, and local district level. Everything seemed to work, all the pieces seemed to be in place, and a high degree of coherence and "ownership" seemed to be maintained. The final piece of the study consisted of observation evaluations of classroom instruction by teachers participating in the program. The teachers were generally strong supporters of the program, but the evaluations showed that most were not following the common curriculum nor adhering to the common standards; each teacher was following his or her own path. However, each teacher thought he or she was sticking with the program. Left on their own to interpret the curriculum their own way, without any outside monitoring, verification, or support, they each went their own way (Cohen 1993; Grant 1993).

Work conducted by National Center for Education Statistics and James Stigler, involving videotapes of 8th-grade classroom instruction in Japan, Germany, and the United States, seconded the conclusion. U.S. teachers think they are implementing curricular reforms, but generally, they are not (U.S. Department of Education 1996, 44-47).

Richard Elmore (1996) reviewed two attempts at large-scale U.S. school reform and, combining his reviews with his readings of the failures of other U.S. curricular reform projects, concluded that schools and their incentive structures are organized in such a way as to thwart reform in curriculum and instruction.

David F. Labaree (1999, 19) offered several compelling reasons for "the chronic failure of curriculum reform":

- Loose coupling of school systems: . . . Administrators have little power to make teachers toe the line instructionally [because they] can fire teachers only with the greatest difficulty, and pay levels are based on years of service and graduate credits, not job performance.
- Adaptability of the school system: . . . Teachers adopt the language and the feel of a reform effort without altering the basic way they do things [and] the differentiation of subjects frees schools to adopt new programs and courses by the simple process of addition. . . . They can always tack on another segment in the already fragmented curriculum [without changing any of the rest].
- Weak link between teaching and learning: . . . Students, after all, are willful actors who learn only what they choose to learn. . . . The law says they have to attend school until they are 16 years old; the job market pressures them to stay in school even longer than that. . . . But these forces guarantee only attendance, not engagement in the learning process.

Note that these three problems either do not exist or are far less potent in highly integrated systems with many enforced quality controls where teachers are evaluated based on actual performance; reforms to a required, core curriculum cannot just be tacked on as an elective; and students have to listen and study if they want to graduate.

It could be, then, that U.S. reforms in the past have faded before they reached the student due to poor quality control in curriculum and instruction systems that were not fully integrated.

APPENDIX
Sources of country-specific information, by country

BELGIUM (FLEMISH COMMUNITY)

Brusselmans-Dehairs, C. 1995, June. *Methods of educational monitoring in the Flemish Community of Belgium*. Ghent, Belgium: University of Ghent.

Dunon, Rita. 1991. *Belgium*, *Special survey on standards and assessments*. National Center for Education Statistics, U.S. Department of Education.

Eurostat. 1995. "Belgium" Structures of the Education Systems in the European Union. EURYDICE/CEDEFOP.

Georis, P., and M. Vilain. 1995. Belgium. In *Handbook of world education*, 81-85.
Monseur, C., and C. Brusselmans-Dehairs. 1997. Belgium. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.
Organisation for Economic Co-operation and Development. 1993. *Belgium: Reviews of national policies for education*. Paris: Organisation for Economic Co-operation and Development.
Philipparat, A. 1996. Belgium. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.
Standaert, R. 1994, September. *The question of national standards in secondary education in the light of Belgium as a federal state*. Brussels, Belgium: Department of Educational Development.
Vanbergen, P. 1988. Belgium. In *The encyclopedia of comparative education and national systems of education*, edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

COLUMBIA

Diaz, C. J., E. Solarte, and J. Arce. 1997. Columbia. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.
Mora, J. 1996. Columbia. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

CYPRUS

Papanastasiou, C. 1996. Cyprus. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.
Papanastasiou, C. 1996. Cyprus. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.

CZECH REPUBLIC

Bishop, J. H. 1997. Do curriculum-based external exit exam systems enhance student achievement? Working paper no. 97-28, Center for Advanced Resource Studies, School of Industrial and Labor Relations, Cornell University, Ithaca, NY.
Kotásek J., and J. Švecová. 1996. Czech Republic. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.
Organisation for Economic Co-operation and Development. 1996. *Czech Republic, reviews of national policies for education*. Paris: Organisation for Economic Co-operation and Development.
Pelavin Research Institute. 1996. *The educational systems of eight countries*. Council for Basic Education.
Petráçek, S. 1988. Czechoslovakia. In *The encyclopedia of comparative education and national systems of education*, edited by T. N. Postlethwaite. Oxford, UK: Pergamon.
Švecová, J., and J. Strakova. 1997. Czech Republic. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.

FRANCE

American Federation of Teachers. 1995. *Defining world class standards*. Washington, DC: American Federation of Teachers.

Bishop, J. H. 1993. Impacts of school organization and signaling on incentives to learn in France, the Netherlands, England, Scotland, and the United States. Working paper no. 93-21, Center for Advanced Resource Studies, School of Industrial and Labor Relations, Cornell University, Ithaca, NY.

Britton, E. D., and S. A. Raizen. 1996. *Examining the examinations: An international comparison of science and mathematics examinations for college-bound students*. Boston: Kluwer Academic.

Eicher, J. C. 1988. France. In *The encyclopedia of comparative education and national systems of education*, edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

Medrich, E. A., S. Kagehiro, and J. Houser. 1994. *Vocational education in G-7 countries: Profiles and data*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Michel, A. 1995. France. In *Performance standards in education: In search of quality*, edited by the Organisation for Economic Co-operation and Development. Paris: A. Michel.

Monchablon, A. 1996. France. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

New Standards Project, Learning Research and Development Center, University of Pittsburgh. 1994, June. *International benchmarking*. Pittsburgh, PA: New Standards Project, Learning Research and Development Center, University of Pittsburgh.

Pelavin Research Institute. 1996. *The educational systems of eight countries*. Council for Basic Education.

Resnick, L. B., K. J. Nolan, and D. P. Resnick. 1995. Benchmarking education standards. *Educational Evaluation and Policy Analysis* 17 (4): 438-61.

Servant, A. 1997. France. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.

Stevenson, H., and S. Lee. 1997. *International comparisons of entrance and exit examinations*. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

GERMANY

Council of Chief State School Officers. 1991. *European lessons from school and the workplace*. Washington, DC: Council of Chief State School Officers.

Eckstein, M. A., and H. J. Noah. 1993. *Secondary school examinations: International perspectives on policies and practice*. New Haven, CT: Yale University Press.

Lehmann, R. H. 1996. Germany. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

Medrich, E. A., et al. 1994. *Vocational education in G-7 countries: Profiles and data*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Riquarts, K. 1997. Germany. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.

Schirp, H. 1995. Germany (North Rhine-Westphalia). In *Performance standards in education: In search of quality*, edited by the Organisation for Economic Co-operation and Development. Paris: H. Schirp.

Stevenson, H., and S. Lee. 1997. *International comparisons of entrance and exit examinations*. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

GREECE

Kontogiannopoulou-Polydorides, G., V. Koulaidis, G. Stamelos, and J. D. Solomon. 1997. Greece. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.

Kontogiannopoulou-Polydorides, G., T. Mylonas, J. Solomon, and D. Vergidis. 1996. Greece. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

Organisation for Economic Co-operation and Development. 1997. *Greece: Reviews of national policies for education*. Paris: Organisation for Economic Co-operation and Development.

ICELAND

Gudmundsson, E., A. Kristjansdottir, and S. Bergman. 1997. Iceland. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.

Larusson, H. 1996. Iceland. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

IRAN

Aziz-Zadeh, H. 1996. Iran. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

Kiamanesh, A. R., and F. Faghihi. 1997. Iran. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.

JAPAN

Bishop, J. H. 1997. Do curriculum-based external exit exam systems enhance student achievement? Working paper no. 97-28, Center for Advanced Resource Studies, School of Industrial and Labor Relations, Cornell University.

Britton, E. D., and S. A. Raizen. 1996. *Examining the examinations: An international comparison of science and mathematics examinations for college-bound students*. Boston: Kluwer Academic.

Eckstein, M. A., and H. J. Noah. 1993. *Secondary school examinations: International perspectives on policies and practice*. New Haven, CT: Yale University Press.

Ellington, L. 1992. Japanese education. Digest no. 5, National Clearinghouse for Japan Studies, Indiana University, May.

Ishino, T. 1995. Japan. In *Performance standards in education: In search of quality*, edited by the Organisation for Economic Co-operation and Development. Paris: T. Ishino.

Johnson, M. L., and J. R. Johnson. 1996. Daily life in Japanese high schools. Digest no. 9, National Clearinghouse for Japan Studies, Indiana University, October.

Kanaya, T. 1988. Japan. In *The encyclopedia of comparative education and national systems of education*, edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

Kanaya, T. 1996. Japan. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

Medrich, E. A., S. Kagehiro, and J. Houser. 1994. *Vocational education in G-7 countries: Profiles and data*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Miyake, M., and E. Nagasaki. 1997. Japan. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.

New Standards Project, Learning Research and Development Center, University of Pittsburgh. 1994, June 11. *International benchmarking*. Pittsburgh, PA: New Standards Project, Learning Research and Development Center, University of Pittsburgh.

Pelavin Research Institute. 1996. *The educational systems of eight countries*. Council for Basic Education.

Stevenson, H., and S. Lee. 1997. *International comparisons of entrance and exit examinations*. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Stevenson, H. W., and J. W. Stigler. 1992. *The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education*. New York: Summit.

U.S. Department of Education, Office of Educational Research and Improvement, OERI Japan Study Team. 1987, January. *Japanese education today.* Washington, DC: U.S. Department of Education.

U.S. Department of Education, Office of Policy and Planning. 1992. *Education standards in the Asia-Pacific Region*. Washington, DC: Office of Policy and Planning, U.S. Department of Education.

KOREA

Bishop, J. H. 1997. Do curriculum-based external exit exam systems enhance student achievement? Working paper no. 97-28, Center for Advanced Resource Studies, School of Industrial and Labor Relations, Cornell University, Ithaca, NY.

Hyde, G.D.M. 1988. *South Korea: Education, culture, and economy.* New York: St. Martin's.

Ihm, C. S. 1993. Restructuring of national education in Korea: New standards of quality, accountability, and vocationalization. In *Education standards for the 21st century: Papers presented at the Asia-Pacific Economic Cooperation Education Symposium*, edited by the U.S. Department of Education, Office of the Under Secretary, Planning and Evaluation Service. Washington, DC: C. S. Ihm.

Kim, J. 1997. Republic of Korea. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.

Lee, K. W. 1983. *Human resource planning in the Republic of Korea: Improving technical education and vocational training*. Washington, DC: World Bank.

McGinn, N. F., and Q. Y. Kim. 1980. *Education and development in Korea.* Cambridge, MA: Harvard.

Shin, S. 1996. Korea, Republic of. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

Sorenson, Clark W. 1994. Success and education in South Korea. *Comparative Education Review* 38(1).

UNESCO Regional Office for Education in Asia and the Pacific. 1992. *New directions in technical and vocational education*. Bangkok, Thailand: UNESCO Regional Office for Education in Asia and the Pacific.

U.S. Department of Education, Office of Policy and Planning. 1992. *Education standards in the Asia-Pacific Region*. Washington, DC: Office of Policy and Planning, U.S. Department of Education.

Yoo, H. J. 1988. Korea, Republic of. In *The encyclopedia of comparative education and national systems of education*, edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

LATVIA

Geske, A. 1997. Latvia. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.

Piebalgs, A. 1996. Latvia. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

LITHUANIA

Liksiene, M., and Z. Jackunas. 1996. Lithuania. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

Zabulionis, A. 1997. Lithuania. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.

THE NETHERLANDS

Bishop, J. H. 1993. Impacts of school organization and signaling on incentives to learn in France, the Netherlands, England, Scotland, and the United States. Working paper no. 93-21, Center for Advanced Resource Studies, School of Industrial and Labor Relations, Cornell University, Ithaca, NY.

Bishop, J. H. 1997. Do curriculum-based external exit exam systems enhance student achievement? Working paper no. 97-28, Center for Advanced Resource Studies, School of Industrial and Labor Relations, Cornell University, Ithaca, NY.

Keersemaker, M. 1991. *Netherlands. Special survey on standards and assessments.* National Center for Education Statistics, U.S. Department of Education.

Kulper, W., and A. Knuver. 1997. Netherlands. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.

Louis, K. S., and Versloot, B. 1996. High standards and cultural diversity: Cautionary tales of comparative research—A comment on "Benchmarking education standards." *Educational Evaluation and Policy Analysis* 18 (3): 253-61.

New Standards Project, Learning Research and Development Center, University of Pittsburgh. 1994, June. *International benchmarking*. Pittsburgh, PA: New Standards Project, Learning Research and Development Center, University of Pittsburgh.

Nijhof, W. J., and Streumer. 1988. Netherlands. In *The encyclopedia of comparative education and national systems of education*, edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

Organisation for Economic Co-operation and Development. 1991. *Netherlands, Reviews of National Policies for Education*. Paris: Organisation for Economic Co-operation and Development.

Resnick, L. B., K. J. Nolan, and D. P. Resnick. 1995. Benchmarking education standards. *Educational Evaluation and Policy Analysis* 17 (4): 438-61.
Vuyk, E. J. 1996. Netherlands. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

PHILIPPINES

Ibe, M. D., and A. E. Punzalen. 1997. Philippines. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.
Postlethwaite, T. N., ed. 1996. *International encyclopedia of national systems of education*, 2d ed. Oxford, UK: Pergamon.

PORTUGAL

Rau, J. M. 1996. Portugal. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

ROMANIA

Noveanu, G. N., V. L. Pop, M. Ludu, and D. Noveanu. 1997. Romania. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.
Vaideanu, G. 1996. Romania. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

SINGAPORE

Chan, S. E., S. T. Chang, and M. Toh. 1997. Singapore. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.
Thomas, R. M. 1988. Singapore. In *The encyclopedia of comparative education and national systems of education*, edited by T. N. Postlethwaite. Oxford, UK: Pergamon.
U.S. Department of Education, Office of Policy and Planning. 1992. *Education standards in the Asia-Pacific region*. Washington, DC: Office of Policy and Planning, U.S. Department of Education.
Yeoh, O. C. 1996. Singapore. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

SPAIN

Gil, G. A. 1996. Spain. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.
Organisation for Economic Co-operation and Development. 1992. *Spain: Reviews of national policies for education*. Paris: Organisation for Economic Co-operation and Development.
Tiana-Ferrer, A. 1995. Spain. In *Performance standards in education: In search of quality*, edited by Organisation for Economic Co-operation and Development. Paris: A. Tiana-Ferrer.

SWITZERLAND

Bishop, J. H. 1997. Do curriculum-based external exit exam systems enhance student achievement? Working paper no. 97-28, Center for Advanced Resource Studies, School of Industrial and Labor Relations, Cornell University, Ithaca, NY.

Gretler, A. 1988. Switzerland. *The encyclopedia of comparative education and national systems of education*, edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

Gretler, A. 1996. Switzerland. In *International encyclopedia of national systems of education*, 2d ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

Moser, U., A. Gretler, E. Ramseier, and P. Labudde. 1997. Switzerland. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.

UNITED STATES

Bishop, J. H. 1993. Impacts of school organization and signaling on incentives to learn in France, the Netherlands, England, Scotland, and the United States. Working paper no. 93-21, Center for Advanced Resource Studies, School of Industrial and Labor Relations, Cornell University, Ithaca, NY.

Bishop, J. H. 1997. Do curriculum-based external exit exam systems enhance student achievement? Working paper no. 97-28, Center for Advanced Resource Studies, School of Industrial and Labor Relations, Cornell University, Ithaca, NY.

Britton, E. D., and S. A. Raizen. 1996. *Examining the examinations: An international comparison of science and mathematics examinations for college-bound students*. Boston: Kluwer Academic.

Eckstein, M. A., and H. J. Noah. 1993. *Secondary school examinations: International perspectives on policies and practice*. New Haven, CT: Yale University Press.

Medrich, E. A., S. Kagehiro, and J. Houser. 1994. *Vocational education in G-7 countries: Profiles and data*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Robeck, E. C. 1997. United States. In *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*, edited by D. F. Robitaille. Vancouver, Canada: Pacific Educational Press.

NOTES

1. To avoid the time and expense of Office of Management and Budget clearance.

2. Remember that this analysis refers to the period 1994-1995. One could argue that the intervening years have seen greater efforts at quality control in much of the United States.

3. It is more likely, however, that when a condition or action is never mentioned, it does not exist.

4. International indicators on this topic seem as confusing as helpful. For example, Tables E5.2b and E5.4a in the Organisation for Economic Co-operation and Development's (OECD's) *Education at a Glance: OECD Indicators 1998*. The TIMSS high performers included seem to defer as much decision making regarding the organization of instruction to the local and school

level as the United States does. Although the percentage of decisions "set within [a] framework set by a higher authority" is not much higher in France and Korea (38%) than it is in the United States (31%), the "framework," however, may be more mandatory and more enforced in France and Korea.

REFERENCES

American Federation of Teachers. 1995a. *Defining world class standards*. Washington, DC: American Federation of Teachers.

American Federation of Teachers. 1995b. *Making standards matter: A fifty-state progress report on efforts to raise academic standards*. Washington, DC: American Federation of Teachers.

Asia-Pacific Economic Cooperation. 1998. *School education statistics in the Asia-Pacific region*. Canberra: Australian Department of Employment, Education, Training, and Youth Affairs.

Beaton, A. E. 1996. *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study.* Chestnut Hill, MA: Boston College.

Beatty, A. 1997. *Taking stock: What have we learned about making education standards internationally competitive? Summary of a workshop*. National Research Council, Commission on Behavioral and Social Sciences and Education, Board on International Comparative Studies in Education. Washington, DC: National Academy Press.

Bishop, J. H. 1993. Impacts of school organization and signaling on incentives to learn in France, the Netherlands, England, Scotland, and the United States. Working paper no. 93-21, Center for Advanced Resource Studies, School of Industrial and Labor Relations, Cornell University, Ithaca, NY.

Bishop, J. H. 1997. Do curriculum-based external exit exam systems enhance student achievement? Working paper no. 97-28, Center for Advanced Resource Studies, School of Industrial and Labor Relations, Cornell University, Ithaca, NY.

Britton, E. D., and S. A. Raizen. 1996. *Examining the examinations: An international comparison of science and mathematics examinations for college-bound students*. Boston: Kluwer Academic.

Brusselmans-Dehairs, C. 1995. *Methods of educational monitoring in the Flemish Community of Belgium*. Ghent, Belgium: University of Ghent.

Cohen, D. 1993. Multiple readings of state instructional policy: Perspectives of state policy makers, district, and school administrators, and teachers. Panel presented at the 15th Annual Research Conference of the Association for Public Policy Analysis and Management, Washington, DC, October.

Council for Basic Education. 1996. An international focus. In *Basic education*. Washington, DC: Council for Basic Education.

Council of Chief State School Officers. 1991. *European lessons from school and the workplace*. Washington, DC: Council of Chief State School Officers.

Cross, C. T., and A. R. Stempel. 1995. Where in the world are we? The need for international benchmarking. Paper presentation at U.S. Department of Education, Office of Educational Research and Improvement, National Institute on Student Achievement, Curriculum, and Assessment, International benchmarking: Setting world class standards to improve student achievement: Working group meeting. Washington, DC.

Eckstein, M. A., and H. J. Noah. 1993. *Secondary school examinations: International perspectives on policies and practice*. New Haven, CT: Yale University Press.

Elmore, R. F. 1996. Getting to scale with good educational practice. *Harvard Educational Review* 66 (1): 1-26.

Fuhrman, S. H. 1993. *Designing coherent education policy: Improving the system*. San Francisco: Jossey-Bass.

Gandal, M. 1997. Leveraging high-performance with gateway examinations. Paper presented at panel "International Benchmarking: New Findings," at the Annual Meeting of the American Educational Research Association, Chicago.

Garet, M. S., and J. Moskowitz. 1996. *Multilateral comparisons of national assessments in education*. Washington, DC: Pelavin Research Institute.

Grant, S. G., 1993. The response of classroom teachers. Paper presented at the 15th Annual Research Conference of the Association for Public Policy Analysis and Management, Washington, DC, October.

Hyde, G.D.M. 1988. *South Korea: Education, culture, and economy.* New York: St. Martin's.

Ihm, C. S. 1993. Restructuring of national education in Korea: New standards of quality, accountability, and vocationalization. In *Education standards for the 21st century*, edited by U.S. Department of Education, Office of the Under Secretary, Planning and Evaluation Service. Papers presented at the Asia-Pacific Economic Cooperation Education Symposium. Washington, DC: C. S. Ihm.

Kreeft, Henk P. J., Ed. 1990. *Issues in public examinations*. CITO, International Association for Education Assessment, June.

Labaree, D. F. 1999. The chronic failure of curriculum reform. *Education Week on the Web*, 19 (May).

Lee, K. W. 1983. *Human resource planning in the Republic of Korea: Improving technical education and vocational training*. Washington, DC: World Bank.

Louis, K. S., and Versloot, B. 1996. High standards and cultural diversity: Cautionary tales of comparative research—A comment on "Benchmarking education standards." *Educational Evaluation and Policy Analysis* 18 (3): 253-61.

Martin, M. O., I.V.S. Mullis, E. J. Gonzalez, T. A. Smith, and D. L. Kelly. 1999. *School contexts for learning and instruction: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

McGinn, N. F., and Q. Y. Kim. 1980. *Education and development in Korea*. Cambridge, MA: Harvard.

McKnight, C. C., Crosswhite, F. J., J. A. Dossey, E. A. Kifer, J. O. Swafford, K. J. Travers, and T. J. Cooney. 1987. The underachieving curriculum: Assessing U.S. school mathematics from an international perspective. Champaign, IL: Stipes.

Medrich, E. A., S. Kagehiro, and J. Houser. 1994. *Vocational education in G-7 countries: Profiles and data*. Washington, DC: National Center for Education Statistics, U.S. Education Department.

Ministry of Education, The Republic of Korea. 1993. *Education in Korea*, *1992-1993*. Seoul: Ministry of Education, The Republic of Korea.

Ministry of the Flemish Community, Department of Education. 1996. Primary education in Flanders core curriculum: Developmental and final objectives, Text of the Government of Flanders, The Act of February 22, 1995, Principles of the Explanatory Statement, October.

Mullis, I.V.S. 1997a. Benchmarking toward world-class standards: Some characteristics of the highest performing school systems in the TIMSS. Paper presented at panel "International Benchmarking: New Findings," at the Annual Meeting of the American Educational Research Association, Chicago.

Mullis, I.V.S. 1997b. *Mathematics achievement in the primary school years: IEA's Third International Mathematics and Science Study.* Chestnut Hill, MA: Boston College.

Mullis, I.V.S., M. O. Martin, A. E. Beaton, E. J. Gonzalez, D. L. Kelly, and T. A. Smith. 1998. *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study.* Chestnut Hill, MA: Boston College.

National Center on Education and the Economy. 1994. *The international experience with school leaving examinations*. Rochester, MN: National Center on Education and the Economy.

National Endowment for the Humanities. 1991. *National tests: What other countries expect their students to know.* Washington, DC: National Endowment for the Humanities.

National Research Council, Center for Science, Mathematics, and Engineering Education. 1996. *Mathematics and science education around the world: What can we learn from the Survey of Mathematics and Science Opportunities (SMSO) and the Third International Mathematics and Science Study (TIMSS)?* Washington, DC: National Academy Press.

New Standards Project, Learning Research and Development Center, University of Pittsburgh. 1994, June. *International benchmarking*. Pittsburgh, PA: New Standards Project, Learning Research and Development Center, University of Pittsburgh.

New Standards Project, Learning Research and Development Center, University of Pittsburgh. 1995. *Performance standards: Draft 5.1*. Pittsburgh, PA: New Standards Project, Learning Research and Development Center, University of Pittsburgh.

Nolan, K. 1997. Implementing world class standards in U.S. state assessment systems. Paper presented at panel "International Benchmarking: New Findings," at the Annual Meeting of the American Educational Research Association, Chicago.

Nolin, M. J., C. Rowand, and E. Farris. 1994. *Public elementary teachers' views on teacher performance evaluations*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Organisation for Economic Co-operation and Development. 1989. *Schools and quality: An international report*. Paris: Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development. 1995a. *Measuring the quality of schools*. Paris: Centre for Educational Research and Innovation, Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development. 1995b. *Performance standards in education: In search of quality.* Paris: Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development. 1995c. *Schools under scrutiny.* Paris: Centre for Educational Research and Innovation, Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development. 2000. *Investing in education: Analysis of the 1999 world education indicators*. Paris: Centre for Educational Research and Innovation, Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development, Indicators of Education Systems Project, Network A. 1995-1996. *Review of assessment activities*. Washington, DC: National Center for Education Statistics.

Peak, L. 1997. *Pursuing excellence: A study of U.S. eighth-grade mathematics and science teaching, learning, curriculum, and achievement in international context*. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Pelavin Research Institute. 1996. *The educational systems of eight countries*. Council for Basic Education.

Phelps, R. P. 1996. Are U.S. students the most heavily tested on earth? *Educational Measurement: Issues and Practice* 15 (3): 19-27.

Phelps, R. P. 2000. Trends in large-scale testing outside the United States. *Educational Measurement: Issues and Practice* 19 (1): 11-21.

Postlethwaite, T. N., ed. 1988. *The encyclopedia of comparative education and national systems of education*. Oxford, UK: Pergamon.

Postlethwaite, T. N., ed. 1996. *International encyclopedia of national systems of education*, 2d ed. Oxford, UK: Pergamon.

Resnick, L. B., Nolan, K. J., and Resnick, D. P. 1995. Benchmarking education standards. *Educational Evaluation and Policy Analysis* 17 (4): 438-61.

Robitaille, D. F., ed. 1997. *National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS*. Vancouver, Canada: Pacific Educational Press.

Schmidt, W. H., D. Jorde, L. S. Cogan, E. Barrier, I. Gonzalo, U. Moser, K. Shimizu, T. Sawada, G. A. Valverde, McKnight, R. S. Prawat, D. Wiley, S. A. Raizen, E. D. Britton, and R. G. Wolfe. 1996a. *Characterizing pedagogical flow: An investigation of mathematics and science teaching in six countries*. Boston: Kluwer Academic.

Schmidt, W. H., C. McKnight, G. A. Valverde, R. T. Houang, and D. E. Wiley. 1996b. *Many visions, many aims: A cross-national investigation of curricular intentions in school mathematics*. Boston: Kluwer Academic.

Schmidt, W. H., C. McKnight, and R. A. Raizen. 1997. *A splintered vision: An investigation of U.S. science and mathematics education*. Boston: Kluwer Academic.

Shanker, A. 1996. *A proposal for a national benchmarking institute on educational standards*. Washington, DC: American Federation of Teachers.

Sorenson, Clark W. 1994. Success and education in South Korea. *Comparative Education Review* 38 (1): 17-29.

Spendolini, M. J. 1992. *The benchmarking book*. New York: American Management Association.

Standaert, R. 1994. *The question of national standards in secondary education in the light of Belgium as a federal state*. Brussels, Belgium: Department of Educational Development.

Stevenson, H., and S. Lee. 1997. *International comparisons of entrance and exit examinations*. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Stevenson, H. W., and J. W. Stigler. 1992. *The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education*. New York: Summit.

UNESCO Regional Office for Education in Asia and the Pacific. 1992. *New directions in technical and vocational education*. Bangkok, Thailand: UNESCO Regional Office for Education in Asia and the Pacific.

U.S. Department of Education, Office of Educational Research and Improvement. 1991. *Improving the math and science curriculum: Choices for state policymakers*. Report on the Secretary's Conference on Improving Mathematics and Science Education.

U.S. Department of Education, Office of Educational Research and Improvement. 1996. *Pursuing excellence: A study of U.S. eighth-grade mathematics and science teaching, learning, curriculum, and achievement in international context* (NCES 97-198). Washington, DC: U.S. Department of Education.

U.S. Department of Education, Office of Educational Research and Improvement, National Institute on Student Achievement, Curriculum, and Assessment. 1995. International benchmarking: Setting world class standards to improve student achievement. Working group meeting, Washington, DC.

U.S. Department of Education, Office of Educational Research and Improvement, OERI Japan Study Team. 1987. *Japanese education today.* Washington, DC: U.S. Department of Education.

U.S. Department of Education, Office of Policy and Planning. 1992. *Education standards in the Asia-Pacific Region*. Washington, DC: Office of Policy and Planning, U.S. Department of Education.

U.S. Department of Education, Office of the Under Secretary, Planning and Evaluation Service. Education standards for the 21st century. Papers presented at the Asia-Pacific Economic Cooperation Education Symposium. Washington, DC: Planning and Evaluation Services, Office of the Under Secretary, U.S. Department of Education.

Yeoh, O. C. 1996. Singapore. In *International encyclopedia of national systems of education*, 2d. ed., edited by T. N. Postlethwaite. Oxford, UK: Pergamon.

Richard P. Phelps has taught secondary school mathematics and science (in Burkina Faso, West Africa) and has worked at the Indiana Education Department, the U.S. General Accounting Office, and the American Institutes for Research. He was the first coordinator of the World Education Indicators Programme at the Organisation for Economic Co-operation and Development in Paris. He now works as senior study director at Westat. Currently, he writes a regular column on student testing policy for the online magazine EducationNews.Org *and, along with Wayne Camara, John Fremer, Steve Ferrara, and Diane Garavaglia, is writing a book,* The War of Testing*, due out this fall from Lawrence Erlbaum.*