DOCUMENT RESUME

ED 425 188		TM 029 252

TITLE	Setting Higher Sights: A Need for More Demanding Assessments for U.S. Eighth Graders.
INSTITUTION	American Federation of Teachers, Washington, DC.
PUB DATE	1998-07-00
NOTE	22p.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Difficulty Level; *Educational Assessment; Foreign Countries; *Grade 8; International Education; Junior High Schools; *Mathematics Achievement; *National Competency Tests; Standards; State Programs; *Test Use; Testing Programs
IDENTIFIERS	Third International Mathematics and Science Study

ABSTRACT

An examination of the content and level of mastery required of students taking statewide mathematics achievement tests was conducted to provide clues about the kind and level of mathematics valued in the United States. Proposals for voluntary national tests also contributed to the rationale for the study. In particular, this study: (1) examined eighth grade mathematics testing by the states to see if, in fact, a de facto national test already existed as a result of widespread use of a limited number of commercial tests; (2) analyzed the content and rigor of three U.S. commercial and two state eighth grade mathematics tests; and (3) compared the content and rigor of U.S. tests to those used by Japan and France, two countries in which students outperformed American eighth graders on the Third International Mathematics and Science Study. Findings indicate: the United States does appear to have a de facto national test in mathematics that is visible in current mathematics test content; tests taken by large percentages of students across the country that are either the same or similar assess low-level content and difficulty at the eighth grade level; and because existing tests determine what gets taught and what materials get published, they cannot move the United States to the goal of being first in the world. (Contains 1 chart, 2 tables, and 15 references.) (SLD)

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

TM

Setting Higher Sights

A Need for More Demanding Assessments for U.S. Eighth Graders

ED 425 188

American Federation of Teachers

TM029252

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

D.H. Gursky

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ERIC

Setting Higher Sights

A Need for More Demanding Assessments for U.S. Eighth Graders

American Federation of Teachers
July 1998

Table of Contents

Executive Summary	4
Introduction	7
The Purpose of this Study	9
What We Did	10
What We Found	14
Conclusions	17
Recommendations	19
Bibliography	20

ERIC
Full Text Provided by ERIC

Executive Summary

During the past 15 months, the National Center for Education Statistics has released three reports on the Third International Mathematics and Science Study (TIMSS). The first report showed United States fourth graders slightly above the international mean in mathematics. The second report indicated that our eighth graders performed slightly below the international mean. The most recent TIMSS report on the performance of students in their final year of secondary school found United States 12th graders are even less competitive internationally, only out-performing students from Cyprus and South Africa. It is only when we look at United States mathematics students who have taken Advanced Placement calculus that we find a group that is competitive with their international peers. Unfortunately, this represents a far lesser percentage of United States students (5 percent) than the comparison groups of other countries' students (an average of 19 percent). It appears that as students progress through the American education system, they fall farther and farther behind in mathematics compared to students from other countries.

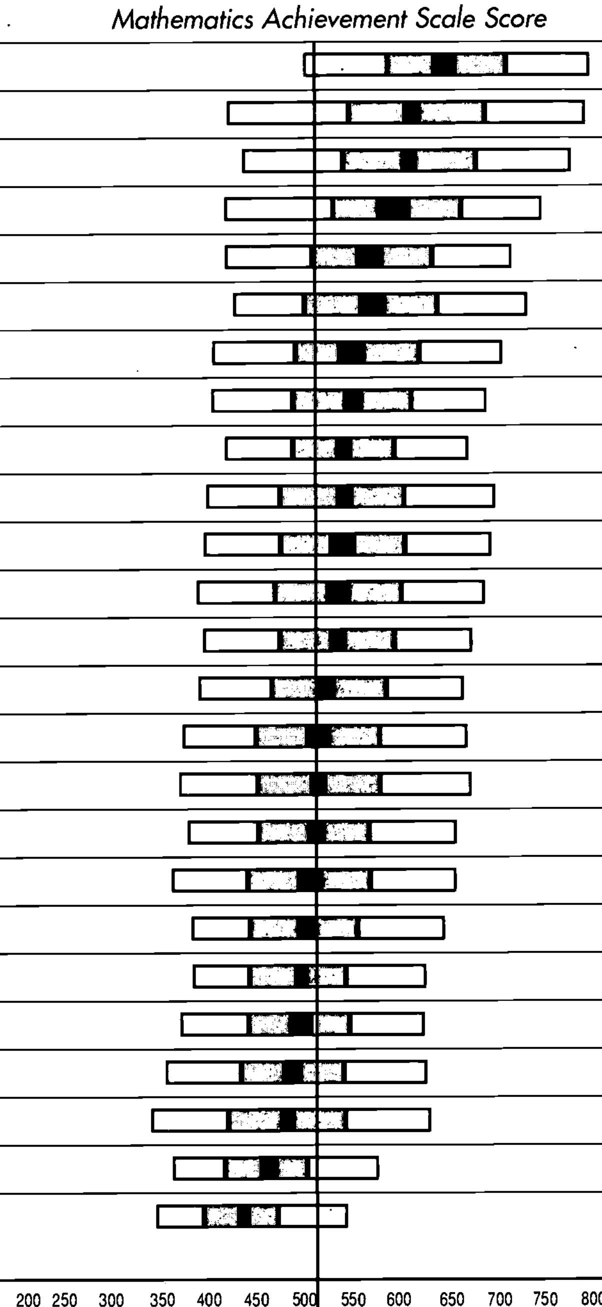
When the end of secondary school report was released, U.S. Secretary of Education Richard Riley rightly said this result was unacceptable. He pointed to the demands for specialized skills in math, science and technology that business and industry are currently unable to fill. Ninety percent of new jobs being created require more than the general knowledge of math and science on which United States students ranked nearly last.

The latest TIMSS report also bursts the myths that the United States is educating more of its students better than most countries and that if we compare only our best against the world's best, American students come out on top. Not only do a greater percentage of other countries' students take higher level mathematics and science courses, but when you compare, for example, all calculus-taking or all physics-taking students, the United States does no better than average—and this with the bloc of high-achieving Asian countries not participating. While the gap between our fifth percentile (lowest-achieving students) and 95th percentile (highest-achieving students) is not larger than most countries, the range of our students' achievement starts and ends at lower levels than is the case for most other countries. For instance, our top quartile performs like the average Japanese or Belgian student, and many of our middle students perform like the bottom quartile of these countries (Table 1).

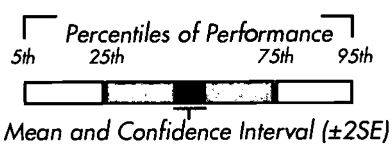
In trying to figure out why United States student performance in mathematics starts out more favorably and then shows a linear downward slide compared to their worldwide peers, one thing we can cite is what the United States expects them to learn. While United States curriculum was defined by TIMSS as extremely broad and not very deep, even in the early grades, at least in those years we target the same content as most of the world, and our students fare well with the basics. But in eighth grade, our students are still studying topics that the rest of the world's students have mastered. In America, mathematics instruction in the middle-school years does not take previously taught content to more complex levels, nor does it introduce challenging material that prepares students for higher-level content

Table 1
Distributions of Mathematics Achievement on TIMSS—Eighth Grade*

Country	Mean	
Singapore	643	(4.9)
Korea	607	(2.4)
Japan	605	(1.9)
Hong Kong	588	(6.5)
Belgium (Fl)	565	(5.7)
Czech Republic	564	(4.9)
Slovak Republic	547	(3.3)
Switzerland	545	(2.8)
France	538	(2.9)
Hungary	537	(3.2)
Russian Federation	535	(5.3)
Ireland	527	(5.1)
Canada	527	(2.4)
Sweden	519	(3.0)
New Zealand	508	(4.5)
England	506	(2.6)
Norway	503	(2.2)
United States	500	(4.6)
Latvia (LSS)	493	(3.1)
Spain	487	(2.0)
Iceland	487	(4.5)
Lithuania	477	(3.5)
Cyprus	474	(1.9)
Portugal	454	(2.5)
Iran, Islamic Rep.	428	(2.2)



Source: IEA Third International Mathematics and Science Study (TIMSS) 1994-95. Beaton, *et. al.*, (1996)



International Average
(Average of all country means = 513)
(Average of 25 countries meeting sampling specifications = 526)

*Only 25 of the TIMSS' 41 countries met all the sampling specifications.

in the eighth grade. This, in turn, affects what happens at grade 12.

It is no secret in America that "what gets tested is what gets taught." Therefore, the AFT reasoned that an examination of what content and level of mastery is required of students taking statewide mathematics achievement tests would provide clues about the kind and level of mathematics that has become valued in the United States. Because American students' low comparative ranking internationally becomes apparent in the eighth grade, we focused our study on eighth-grade mathematics. In addition, because President Clinton, partly in response to the poor eighth-grade showing on TIMSS, proposed that there be a voluntary national test of mathematics achievement of eighth-grade American students benchmarked to international standards, we decided to examine the content of tests commonly used by states and districts to determine the proficiency of their students and to benchmark those tests to tests used in high-performing TIMSS countries.

In particular, our study:

- examined eighth-grade mathematics testing by the states to determine if, in fact, we already have a *de facto* "national test" that is the result of widespread use of a limited number of commercial tests;
- analyzed the content and rigor of United States commercial and state eighth-grade mathematics tests; and
- compared the content and rigor of United States tests to those used by Japan and France, two countries whose students outperformed American eighth graders on the TIMSS.

What did we find?

- The United States does appear to have a *de facto* national test in mathematics that is visible in current mathematics test content. Large percentages of students across the country take the same or similar tests of math achievement.
- Those tests assess low-level content and difficulty at the eighth-grade level.
- Existing tests are incapable of providing information about high-end performance because such performance is not tested.
- Since existing tests drive what gets taught and what mathematics materials get published, they cannot move us to achieve our goal of being first in the world.

These findings indicate that we need a national, voluntary test that—unlike current *de facto* national tests—pushes us to make progress toward meeting the world-class standards that students reach in high-achieving TIMSS countries. But it is critical to understand that tests alone, no matter how intellectually demanding and how consequential to the lives of students, will not, in themselves, yield higher achievement of our youth. A well-developed, highly focused mathematics curriculum must be in place, and teachers must be prepared, in terms of both pedagogy and mathematics content, to assist students in mastering new, complex material.

ERIC

Introduction

In the fall of 1996, the United States released a report, *Pursuing Excellence*, which compared United States eighth-grade mathematics achievement with that of eighth graders in 41 other countries, including several of our major international competitors—Canada, England, France, Germany, Italy and Japan. The results of that Third International Mathematics and Science Study (TIMSS) indicated that United States students scored below the international mean for mathematics and were in the middle group of countries overall.[1] Students in Singapore, Korea, Japan, Hong Kong, Belgium, the Czech and Slovak Republics, Switzerland, the Netherlands, Slovenia, Bulgaria, France, Hungary, Russia, Australia, Ireland and Canada all outperformed our eighth graders. Indeed, even our best students were not competitive—only 5 percent of American eighth graders contributed to the top 10 percent of students internationally, as compared with 32 percent of the Japanese students. Clearly, we are far away from our national goal of being first in the world in achievement in math.

But the TIMSS study was more than a "horse race" among nations; it examined other aspects of schooling, including textbooks and curriculum used in the various countries in the international study. The study concluded that the mathematical content of United States lessons in comparison to that of other countries—both in textbooks and in actual practice—is less advanced. American eighth-grade students are still spending considerable time on whole-number computation and fractions and decimals, when most other countries have a strong focus on algebra and geometry.

Partly in response to our mediocre performance in mathematics on the TIMSS assessment, in 1997 President Clinton proposed that the federal government develop voluntary, national tests in fourth-grade reading and eighth-grade math. The tests are to be designed to produce individual scores that can be reported to parents and school officials. Such a national, voluntary assessment will make it possible for parents to compare the performance of their children to national standards (and, in mathematics, to international standards). This new federal testing initiative is

[1] TIMSS also assessed science achievement, but this study is focused on mathematics achievement, so we are not reporting the science findings.

to be based on the fourth-grade reading and eighth-grade National Assessment of Educational Progress (NAEP) frameworks and is to be designed so that individual student scores can be related to NAEP performance levels of "basic," "advanced" and "proficient." In the case of eighth-grade mathematics, predictions also will be made from individual student scores to likely performance on the TIMSS.

The proposed national test addresses a long-standing concern in this country that parents and, indeed, some educators, do not have realistic notions and reliable information on how well their children are doing compared to other students in the nation or to children in other countries. The "Lake Woebegone" effect of many state assessments—that is, all the students are above average—has led policy makers and others to distrust state information on achievement. This distrust is further buttressed by research that shows considerable difference between the scores that students receive on the state NAEP compared to their performance on other state assessments. The state assessments are much more likely to show that students generally are proceeding satisfactorily than are the NAEP results (Musik, 1996).[2]

The NAEP assessments are constructed on the basis of a national consensus process focused on what students in a particular subject area and at a particular grade level should know. NAEP does produce measures of how well students collectively in a state and in the nation perform. It is the only measure that can allow groups of students in all regions of the country to be compared on the same test. But NAEP does not generate individual scores, so parents cannot get an accurate assessment of their child's performance from those tests. President Clinton's voluntary testing proposal is designed to address this problem and to provide parents with accurate, reliable information concerning their child's mathematics performance in comparison to other students in the United States and to that of their international peers.

[2] There has, however, been much serious criticism of NAEP achievement setting procedures and the interpretations about student performance that stem from them. Indeed, NAEP may understate achievement performance of students, thus in part explaining the state-NAEP discrepancies. For an excellent analysis of some of the difficulties see for example, Linn (1998).

ERIC

The Purpose of this Study

While TIMSS results show that the mathematics of classroom instruction and of textbooks used in the United States are not as challenging as that of other countries, TIMSS did not look at testing practices in those countries. We speculated that our poor international showing in mathematics may be in part a result of American assessment practices. It is well known that "what gets tested is what gets taught."

Furthermore, expectations concerning what content is to be mastered at what level in the current state and commercial tests may account for much of the performance differential seen in NAEP/state comparisons and in comparisons of American students and many of their Asian and European peers. For example, the performance differentials noted above may be explained by:

- differences in how intellectually challenging the items are among NAEP, state assessments, commercial tests and international assessments.
- American testing practices that judge the rigor of test items for inclusion in an assessment by the probability of how many students can pass them, given today's undemanding curriculum, rather than by the intellectual demands of the mathematics being assessed.
- political pressure that encourages states to use tests with content or performance levels that make state education efforts and student performance look good.
- curricula that do not touch upon content assessed in NAEP and international assessments; they therefore assess students on topics that they were not taught at the time the test was administered.

Concern about the quality and rigor of current state efforts at mathematics assessment led us to look into this matter. To help inform the debate surrounding the creation of a national, voluntary eighth-grade mathematics test, and to provide information to test developers, particularly those charged with developing state and national assessments, the AFT:

- examined eighth-grade mathematics testing in the states to determine if, in fact, we already have a *de facto* national test that is the result of widespread use of a small number of commercial tests;
- analyzed the content and rigor of our commercial and state eighth-grade mathematics tests; and
- compared the content and rigor of American tests to those used by Japan and France, two countries whose students outperformed United States eighth graders on TIMSS.

We believe that our examination of widely used American tests can help clarify the folly of giving into pressures—political or curricular—to provide students with instant pseudo-success instead of providing them with incentives to match the knowledge levels of their international peers.

ERIC Full Text Provided by ERIC

What We Did

The Tests

The content and context of tests play a crucial role in influencing what both teachers and students come to believe are the implicit goals of mathematics education. Using the Council of Chief State School Officers and North Central Regional Educational Laboratory (NCREL) data bases (CCSSO, 1997), we looked at what tests states use to assess their students' eighth-grade mathematics knowledge and skill. While many states, particularly the larger ones (e.g., New York, Maryland, Texas), use tests that they develop, more than a third of students across the country are tested using commercial tests developed by CTB McGraw Hill, Harcourt Brace and Psychological Corporation. We realized that by examining a handful of tests, we could make a statement about the rigor and quality of assessments that were administered to more than 40 percent of the students in America and that, therefore, greatly influenced the mathematics taught to eighth graders across the country.

For this study, we analyzed eighth-grade tests from three commercial publishers and two states:

1. *Terra Nova*, CTB McGraw Hill;
2. *New Standards*, Harcourt Brace;
3. *Stanford 9*, the Psychological Corporation;
4. The *Texas Assessment of Academic Skills (TAAS)*, Texas Education Agency; and
5. The *New York State Goals 2000 New Assessments Project Test for Grade 8* (piloted in Winter of 1997).

The commercial tests we chose to study were the most recent and most reflective of current products in assessment available on the commercial market. We selected the New York and Texas tests because large numbers of students are assessed by them and because they reflect expectations regarding mathematics achievement for different geographical regions of the country and distinct, but varied, groups of students within each of those areas.

After examining the United States eighth-grade tests, we thought it appropriate to look at tests given in countries whose students outperformed ours. After all, the national test was prompted in part by our mediocre showing on TIMSS at the eighth-grade level, and we have a national goal to be first in the world in mathematics.

We looked at testing in France and Japan. We chose France because, not only do her students always do well, but also the difference in achievement between her best and worst students is one of the smallest in the world. France is able to bring more of her students to higher levels than we do.[3] We chose Japan because students are heterogeneously grouped through grade eight, her students do very well, and there is a good deal of information about the Japanese system of instruction, including some translations of grade seven, eight and nine textbooks. Unlike the United States, where eighth grade is the most popular grade for statewide testing of mathematics and one of the grade levels used for reporting NAEP results, France and Japan conduct their nationwide student assessments of student mathematics achievement at the ninth grade.

Analysis of the Tests

Because President Clinton has indicated that the new national tests will be based on the NAEP content frameworks, we used those frameworks to evaluate the tests. We convened a panel of four individuals—Deborah Paulson, who is an AFT member, Thinking Math Trainer, a Milken Award winner, and mathematics teacher in Texas; John Dossey, former president of the National Council of Teachers of Mathematics, consultant to NAEP and TIMSS, and mathematics professor at Illinois State University; Gail Burrill, current NCTM president and AFT high school mathematics teacher in Wisconsin; and Norman Webb, mathematics professor at the University of Wisconsin and senior scientist at the Wisconsin Center for Education Research—who independently examined the commercial tests, the two state tests and the two international exams.

The panel used the NAEP frameworks to classify the content of the tests they examined. The NAEP content frameworks identify five major topics at that particular grade level: Number (Arithmetic); Measurement; Geometry; Data and Statistics; and Algebra and Functions. Algebra and Functions is divided into 14 subtopics, each of which is specified further. For example, one of the subtopics under Algebra and Functions is "solve systems of equations and inequalities." This has more detailed specifications—"a) solve systems graphically, b) solve systems algebraically and c) solve systems using matrices."

Panelists assigned items on the various tests to one of the five major topics that comprise the NAEP content frameworks. Any differences in assignment among panelists were resolved through discussion and mediation. The differences were few and generally occurred only on items that required knowledge and skills from more than one area of mathematics or that could be solved in more than one way. These items were often classified as more difficult because students were combining knowledge from different areas. Three of our consultants had done such analyses before on the NAEP math items as consultants to NAEP.

Panelists also rated each item's difficulty level. There was little disagreement about the levels assigned to each item, and consensus was reached on 99 percent of all items on all the tests. Panelists used the following definitions of easy, middle level and hard.

- **Easy** items basically require students to recognize and plug numbers into a formula, which is usually given. The solution jumps out at the student. Any "context" is window dressing and is unnecessary to solving the problem. The student can complete the item without having to know relationships or put together information.
- **Middle-level** items require the student to formulate a solution plan. They require thinking, the coalescing of knowledge. They usually require the student to produce some additional information before the final solution, and many require some generalization.
- **Hard** problems require the creation of an abstract model. Understanding the problem, and what it requires, demands effort. The problem is not well defined in that one cannot look at it and immediately know what to do. The context is meaningful and necessary to solving the problem. A student may have to establish a procedure. Arithmetic is not as visible as it is in easy and middle-level problems. The student draws upon logic, theory and proven principles. A hard problem requires students to know

[3] Sixty-three percent of the French students tested in the eighth grade TIMSS scored above the international mean for all 41 countries. (Beaton *et. al.*, 1996)

which theorem is relevant and how to apply it. (Chart 1 presents examples of NAEP items at each of these levels.)

Please note, the definitions refer to what a student must do to solve the problem, to the level of work that must be done by the solver. If all a student must do is look at a formula and plug in the numbers, or if the problem tells the students what to do to solve it, the problem demands only simple skills and is classified as easy. If the solver must put information together and make a plan to solve the problem, if the information that must be used is not all given and the student must create some of it him/herself, the problem is at a high-

Chart 1:
Examples of Easy, Middle Level, and Hard Eighth-Grade Items

Easy items basically require students to recognize and plug numbers into a formula, which is usually given. The solution jumps out at the student. Any "context" is window dressing and is unnecessary to solving the problem. You can complete the item without having to know relationships or put together information.

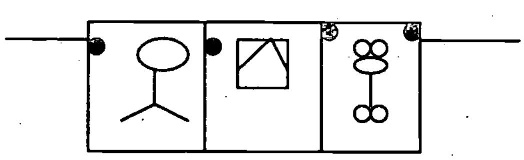
EASY

In a bag of marbles, 1/2 are red, 1/4 are blue, 1/6 are green and 1/12 are yellow. If a marble is taken from the bag without looking, it is most likely to be

A. red.
B. blue.
C. green.
D. yellow.

(73 percent of eighth graders got this right.)

Middle-level items require the student to formulate a solution plan. They require thinking and the coalescing of knowledge. They usually require the solver to produce some additional information before the final solution, and many require some generalization.



MIDDLE LEVEL

Children's pictures are to be hung in a line as shown in the figure above. Pictures that are hung next to each other share a tack. How many tacks are needed to hang 28 pictures in this way?

A. 27.
B. 28.
C. 29.
D. 56

(25 percent of eighth graders got this correct.)

BEST COPY AVAILABLE

er level of difficulty. The definitions represent a consensus of the panel as members responded to our request to describe how to assess the rigor of items on a math test.

Format and Time

Finally, we collected information on the amount of time allotted students to complete the examination, the number of "scoreable units" within the tests (some questions had several parts, each of which was treated as a "scoreable unit") and the percent of items that were multiple choice and open response.

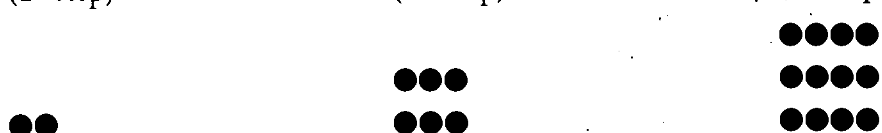
Hard problems require the creation of an abstract model. Understanding the problem and what it requires of you means work. The problem is not well defined in that you cannot look at it and immediately know what to do. The context is meaningful and necessary to solving the problem. You may have to establish a procedure. Arithmetic is not as visible as it is in easy and middle-level problems. The solver draws upon logic, theory and proven principles. A hard problem requires you to know which theorem is relevant and how to apply it.

HARD

This question requires you to show your work and explain your reasoning. You may use drawings, words and numbers in your explanation. Your answer must be clear enough so that another person could read it and understand your thinking. It is important that you show all your work.

A pattern of dots is shown below. At each step, more dots are added to the pattern. The number of dots added at each step is more than the number added in the previous step. The pattern continues infinitely.

(1st step) (2nd step) (3rd step)



Marcy has to determine the number of dots in the 20th step, but she does not want to draw all 20 pictures and then count the dots.

Explain or show how she could do this and give the answer that Marcy would get for the number of dots.

(63 percent of eighth graders got this wrong; 16 percent didn't try it; only 6 percent were credited with either satisfactory or better responses.)

What We Found

The results of the panelists' analyses are presented in Tables 2 and 3. A quick glance at the tables reveals that:

- state and commercial tests are easier and require less-advanced problem solving than the international tests;
- state and commercial tests assess more arithmetic and measurement than do the international tests;
- international tests are more rigorous in regard to the assessment of algebra and geometry than are the United States' tests; and
- items on the international tests are open response and require that students show how they solve problems, whereas the United States tests are predominantly multiple-choice items with little intellectual demand associated with determining the answer.

Content Coverage

Commercial tests—Table 2 contains the results of the analyses of the eighth-grade assessments. The data related to the three commercial standardized eighth-grade tests are in the first three columns. There were clear differences in the distribution of scoreable items across the NAEP content areas among the tests and compared to the distributions on the NAEP assessment. The *New Standards* test and the eighth-grade *Terra Nova* gave higher allotments to the number and operation strand than did the *Stanford 9*, and the *New Standards* test compensated by less testing of geometry and algebraic concepts and skills.[4]

The *Stanford 9* had a notably higher percentage of algebra items at eighth grade than did the other United States tests. *New Standards* focused heavily on data and statistics, one of the stronger areas for United States eighth-grade students on TIMSS. (In the eighth-grade NAEP, 20 percent of the items are devoted to geometry and 25 percent to algebra—45 percent to these two topics.) This compares to a range among the United States commercial and state tests from a low of 10 percent (5 percent each for algebra and geometry—*New Standards*) to a high of 30 percent (10 percent to geometry and 20 percent to algebra—*Stanford* 9).

Because the foreign tests were administered at ninth grade, the panel looked at the ninth-grade *Terra Nova* test as well. The eighth-grade analysis had showed that *Terra Nova* had the highest percentage of items at the middle level of difficulty. Analysis of the ninth-grade *Terra Nova* (see Table 3) indicates some evidence of a shift in content from number and operation to algebraic concepts and skills. This is probably a reflection of the fact that only 24 percent of the nation's youth

[4] All test developers had an opportunity to respond to the results of our analysis of their tests. It should be noted that Phil Daro, the director of development of the New Standards mathematics test, took exception to our analysis because we did not weight the scoreable items. While scaling or forming composite scores could change the weight on an individual item, the analyses in this work were made on the basis of the stimuli students saw as they took the exam. For example, New Standards claims that 20 percent of their test is geometry. This claim is based on the weight they give the few geometry items in the final score, not on the proportion of the 26 scoreable events on their test that assess geometry. According to their weighted analysis of content, the New Standards test is 20 percent number sense, properties and operations, 15 percent measurement, 20 percent geometry and spatial sense, 15 percent data analysis, statistics and probability, and 30 percent algebra and functions. (Personal communication, Phil Daro, June 25, 1997, Washington, D.C.).

Table 2
Analysis of U.S. Eighth-Grade Mathematics Assessments

	New Standards Grade 8	Stanford 9 Grade 8	Terra Nova Grade 8	N.Y. Grade 8	Texas Grade 8
Topics tested					
Number	45%	34%	44%	37%	52%
Measurement	15%	16%	12%	24%	10%
Geometry	5%	10%	12%	5%	5%
Data/Probability	30%	20%	20%	18%	10%
Algebra	5%	20%	12%	16%	23%
Item Difficulty					
Hard	0%	0%	0%	0%	0%
Medium	10%	10%	13%	21%	2%
Easy	90%	90%	87%	79%	98%
Type of response required					
Multiple Choice	0%	85%	73%	0%	100%
Open Response	100%	15%	27%	100%	0%

take Algebra I in the eighth grade; the majority take it as part of their ninth-grade program. Hence, most tests delay their coverage until that grade level. But, even with the increase in attention to algebraic content in the ninth grade, the *Terra Nova* exam contains much lower percentages of algebra and geometry than the foreign tests.

State tests—The two state tests were remarkably different from one another in format; the New York test was all constructed response, while the Texas assessment was all multiple choice. The New York test consisted of 18 application items and one extended item. In contrast, the *TAAS* had 60 multiple-choice items.

But, both tests gave short shrift to the content area of geometry, each allocating only 5 percent of the scoreable items to it. Analysis of the remaining content categories showed that the New York examination had a more even distribution of emphasis across the remaining categories, with number receiving a slightly heavier emphasis than the other areas. The Texas examination, on the other hand, gave only 10 percent of its emphasis to measurement and 10 percent to data, double this to algebra, and 52 percent of the examination's emphasis to number and operation.

Finally, the *TAAS* had the highest percentage of easy items across all the tests we examined, while the New York assessment had the lowest percentage of easy items across all the United States tests.

Foreign tests—In making these comparisons, one must remember that only one of the United States examinations was intended for the ninth grade. But, even in this case, the comparison is stark. Unlike the American tests, which are predominantly multiple choice (with the exception of *New Standards*), the international tests are virtually all constructed response. Furthermore, the tests we looked at—the French *brevet* and the Japanese examinations—focus heavily on number, geometry and algebra, giving little or no emphasis to measurement or data analysis. These tests gave less emphasis to number than did the United States examination, a median of 22 percent of international scoreable items compared to the United States

ERIC

median of 31 percent. In contrast, they each devote about 75 percent of the test to algebra and geometry compared to the United States, which devoted 37 percent. The medial comparisons in geometry are 28 percent on the international tests compared to the United States median of 14 percent. In algebra, the medial comparisons are 47 percent internationally to the United States median of 23 percent.

Problem Difficulty

The story is much the same in the categorization of problems relative to the difficulty level for eighth-grade students. For the commercial and state-developed examinations, from 2 percent to 21 percent of the items were judged of medium difficulty, with the remaining being judged easy (Table 2). A further analysis of the items showed that most were straightforward tasks requiring little or no real information processing on the students' part. This is also reflected in the relationship between the number of items and the time allotted for the students to complete the examination—generally, the American students get more problems and less time to complete them.

Only half of the items on the foreign examinations, on the other hand, were deemed to be at the easy level, as judged by the panelists' definition of rigor (Table 3). It should be noted, however, that *Terra Nova* representatives objected to our judgmental definitions of easy, medium and hard. They use "p values" to determine difficulty level—that is, the percent of students who answer an item correctly, rather than the intellectual demand required to solve the problem. In addition, Daro, the developer of the *New Standards* assessment, had a slightly different take on our definition of difficulty. He asserted that *New Standards* assigns difficulty evaluations not to the items but to the curriculum. For example, in pointing to a probability item, which met our criteria of easy—no new information had to be brought to the table to solve the problem—Daro indicated that it was "only easy if you have been taught the material." Therein lies the dilemma of "p values" and expectations, and it is precisely why our analysis looked at intellectual demand, not student performance, in determining the difficulty level of items.

Table 3
Analysis of U.S. and International Ninth-Grade Mathematics Assessments

	Terra Nova Grade 9	France Grade 9	Japan Grade 9
Item topics			
Number	31%	25%	18%
Measurement	11%	4%	2%
Geometry	14%	25%	30%
Data/Probability	20%	0%	2%
Algebra	23%	46%	48%
Item Difficulty			
Hard	0%	8%	20%
Medium	11%	38%	25%
Easy	89%	54%	50%
Type of response required			
Multiple Choice	71%	0%	5%
Open Response	29%	100%	95%

Conclusions

What conclusions can be drawn from this analysis?

First, given these American tests, our students' performance on the eighth-grade TIMSS assessment is not surprising. American tests have a very different allocation of items to content than do the two foreign tests examined. The difference goes deeper, however, than the topics tested. The tests also focus on the mathematics in different ways. There are very different demands made on students. The foreign tests require students to figure out relationships and strategies, while United States test items remain mostly at a surface level and make few intellectual demands on students.

- It is important to remember that teachers tend to provide the kind of instruction that prepares students for the level and the mathematics that is tested. If tests require students to plug in a formula, complete a rule, or select a predetermined strategy, instruction will focus on procedures, and emphasis will be on one acceptable way to solve a problem. Students will not be encouraged to look for relationships or figure out more than one way to look at problems. Students will be taught how to do particular problems. The knowledge may not be transferable to new situations.
- If, on the other hand, the test includes items where the solution procedure is not apparent, where the solver must organize information before proceeding, where a problem can be solved in more than one way, students will have a chance to tackle problems in various ways. Instruction will stress fundamental mathematical principles, and emphasis will be placed on what and why as well as how.[5]

Second, the tests developed for assessing mathematics performance in the United States, whether they are open response or multiple choice in nature, tend to be overwhelmingly tests of arithmetic and low-level skills. While it is probably necessary to have some items that assess basic arithmetic, this area should not comprise the greatest part of a test at the eighth-grade level. Nonetheless, we found that:

- the commercial and state tests tend to place a heavy emphasis on backtracking to arithmetic skills rather than moving forward to derive a balance between algebraic and geometric reasoning at the eighth-grade level;

[5] This is precisely what Stigler *et. al.*, (1997) found in their international video study of eighth-grade mathematics instruction.

- the commercial and state tests tend to present problems that are already fully formulated and the appropriate numbers are highlighted, ready for students to insert them into appropriate operations. They make little demand on students to interpret and establish relationships reflecting real understanding of the underlying mathematics;
- in other countries, the application items tend to be presented in a rich context and involve both graphic and tabular data, which students are expected to analyze and understand. This context is then drawn upon for a series of independent but related questions. This allows students to think more deeply about a mathematical situation and then exhibit a broader range of abilities to reckon with the situation from a number of vantage points;
- the depth of mathematics required in the United States examinations also differed beyond the actual percentage counts reflected in Tables 2 and 3. While an American and foreign test might have the same percentage of items in a given area, our tests do not make the same demands on student understanding. The American tests tend to provide information and ask students to plug in the information in a formula or well-known computational procedure, while the foreign examinations expect more original work from students. This is especially true in algebra and spatial relations. Students in other countries are expected to create equations and describe three-dimensional cross-sections, while our students are expected to evaluate a simple expression for a whole number substitution or find the area of a rectangle, given its length and width.

In sum, we found that:

- the United States does appear to have a *de facto* national mathematics test in that large percentages of students across the country take the same or similar tests of math achievement;
- those tests assess low-level content and difficulty at the eighth-grade level;
- existing tests are incapable of providing information about high-end performance because it is not tested;
- since existing tests drive what gets taught and the materials that are published, they cannot move us to achieve our goal of being first in the world;
- therefore, we need a national, voluntary test that—unlike current *de facto* national tests—pushes us to make progress toward a world-class standard as shown in TIMSS.

ERIC

Recommendations

This analysis of mathematics tests provides a basis for recommendations regarding the nature and focus for the proposed national, voluntary eighth-grade examinations in mathematics.

- First, because we want the national test to reflect international standards, the test needs to target much more algebra and geometry than TIMSS research show currently contained in our curriculum and that are currently tested on NAEP at the eighth-grade level.

- Second, because it is important to push for rigor in the field of mathematics, the items on the national test should span the entire range of intellectual levels with no more than 50 percent of the scoreable items in the "easy" or lower level. In the other 50 percent, the items in context should require multiple steps for solution. Items should not tell students how to solve problems but require them to figure out what mathematics or strategy is appropriate. Contextual items should not tell students which formula to use and give them that formula. They should provide contexts that require thoughtful use of mathematical knowledge and skill. There should also be items that require students to demonstrate or articulate their understanding of core mathematical principles and properties that will enable students to apply mathematics in new and more complex situations.

- Third, because we know that the students in eighth grade will have a range of achievement levels, and the idea of assessment is to find out what students know, not just what they don't know, there should be problems that can be solved with more and less sophisticated methods or that have multiple parts, some of which could receive partial credit.

- Fourth, the analysis showed that one of the characteristics of United States tests is that they provide less time for students to take them. Since solving complex problems requires more time, this may, in part, account for the low-level items on United States tests. Thus, we recommend that more time be allotted to support the inclusion of more than one complex problem that allows students to demonstrate high-level mathematical thinking.

Critics of the proposed national test argue that we can get the same information such a test would provide by using existing commercial and state tests for eighth-grade students. Indeed, if the national test ends up mirroring the current low expectations of commercial and state tests, the critics will be right, and we will miss a rare and powerful opportunity to raise the standard of achievement expected of United States students. Teachers will continue to teach what is currently the content on low-level tests.

As the national examinations are being developed, it is imperative that their construction, both in item format and intellectual demand, be such that they support efforts to make American mathematics education among the best in the world. When tests expect too little, or test in inappropriate ways, the wrong message is sent to both students and teachers in the nation's classrooms. If done right, however, the tests associated with President Clinton's initiative can have a positive impact on mathematics education and can enable our students to compete with the best in the world, rather than with the mediocre.

Bibliography

American Federation of Teachers, National Center for Improving Science Education. (1997) *Defining World Class Standards: What Students Abroad Are Expected To Know about Mathematics.* "1993 Brevet Exam in Mathematics. Caen Region, France." Washington, DC: AFT. (pp. 8-12).

American Federation of Teachers, National Center for Improving Science Education. (1997) *Defining World Class Standards: What Students Abroad Are Expected To Know about Mathematics.* "1995 High School Entrance Exam in Mathematics. From Tokyo Prefecture." Washington, DC: AFT. (pp. 50-54).

Beaton, A., Mullis, I.V., Martin, M., Gonzalez, E., Kelly, D. & Smith, T. (1996). Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study. Boston: Center for the Study of Testing, Evaluation, and Educational Policy.

Burrill, G., Dossey, J., Paulson, D. & Webb. N. (1997). *AFT Panel on 8th-Grade Assessment.* Washington, DC: American Federation of Teachers.

CTB/McGraw Hill. (1997). *Terra Nova Multiple Assessments: Parts 1 and 2. Grades 8 and 9.* Monterey, CA: CTB/McGraw Hill.

Daro, P. (1997). Personal Conversation. Washington, DC

Linn, R. (1998) Validating Inferences from National Assessment of Educational Achievement-Level Reporting. *Applied Measurement in Education*, 11 (1), 23—47.

Musick (1996). *Setting Education Standards Higher.* Atlanta, GA: Southern Education Regional Board.

New Standards Project: *New Standards Grade 8 Mathematics Assessment.* San Antonio, TX: Psychological Corporation.

New York Department of Education. (1997) *New York State Goals 2000 New Assessment Project: Mathematics.* Grade 8. Albany, NY: New York State Department of Education.

Harcourt-Brace (1995). *Stanford 9 Grade 8 Mathematics Assessment.* San Antonio, TX: Psychological Corporation.

Stigler, J.W., Gonzales, J., Kawanaka, T., Knoll, S. & Serrano, A. (1996). The TIMSS Videotape Classroom Study. Los Angeles: UCLA Psychology Department.

Texas Education Agency. (1996) *Texas Assessment of Academic Skills (TAAS).* Texas Education Agency.

Third International Mathematics and Science Study (1997). *Pursuing Excellence.* Washington, DC: National Center for Education Statistics.

The College Board. (1996) *Mathematics Framework for the 1996 National Assessment of Educational Progress.* Washington, DC: National Assessment Governing Board.

ERIC

AMERICAN FEDERATION OF TEACHERS
AFL-CIO

555 NEW JERSEY AVENUE, NW
WASHINGTON, DC 20001-2079

ASK FOR ITEM NO. 268
7/98

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC

TM029252

NOTICE

REPRODUCTION BASIS

☑ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

(9/92)