

**ADDRESS TO THE 1998 CONFERENCE OF THE NATIONAL
SCIENCE TEACHERS' ASSOCIATION**

**WHAT WE'VE LEARNED FROM TIMSS
ABOUT SCIENCE EDUCATION
IN THE UNITED STATES**

By:

Pascal D. Forgione, Jr., Ph.D.
U.S. Commissioner of Education Statistics
National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, N.W., Room 400
Washington, DC 20208
202-219-1828 (Telephone)
202-219-1736 (Fax)
Pascal_Forgione@ed.gov (email)

NCES World Wide Web Home Page: <http://www.nces.ed.gov>

April 16, 1998

I. Introduction

Let me begin by saying how excited I am to address this group, a group of people who are both teachers and scientists. As the study I will be speaking about is a scientific study of science achievement and education, I hope you find what I have to say interesting and useful, and I look forward to your reaction.

In my talk today I would like to present the findings of the most recently released results from TIMSS, the Third International Mathematics and Science Study. These results, released a couple of months ago in February, cover students at the end of high school. I would then like to briefly respond to some of the criticisms that have been made regarding this assessment. Then, I would like to take a broader look at what TIMSS tells us about science achievement across the levels of schooling, bringing in as appropriate the previously released findings from other components, primarily the fourth- and eighth-grade assessments. I will take a similarly broad look to see if TIMSS offers any explanation for the patterns of our performance. I would then like to highlight some of the questions raised by TIMSS and close by briefly outlining future activities related to international comparisons of science education.

Before I begin, however, I thought it would be helpful to give a brief description of my agency, its roles and responsibilities, and our involvement in the TIMSS project.

The National Center for Education Statistics—or “NCES”—is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to: collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; and review and report on international education activities. Many of you are probably familiar with one of our other major activities, the National Assessment of Educational Progress—or “NAEP”—an assessment of student achievement in various subject areas which has been conducted since 1969. Our major annual publications include *The Condition of Education* and the *Digest of Education Statistics*.

As you might suspect, NCES maintains a virtual sea of data on most any subject related to education. But we do not collect data solely for the sake of having it. Our activities are driven by the responsibility to meet the needs of our various audiences, among whom are legislators, policymakers, researchers, the media, the general public, and most certainly you, teachers. As such, one of our primary goals is to ensure that the data we produce are **useful**, in that they can answer important questions relevant to major decisions of education policy, programs, and practice. To be useful, data must also be presented in a variety of formats. Our publications include reports, books, newsletters, and issue briefs that range in size, level of detail, and scope. The development of the Internet has allowed us to expand our options for presenting data while simultaneously improving our customers’ ability to access it.

In addition to usefulness, two other goals I have set for myself and for NCES are that the data we provide be **regular** and **timely**. By regular, I mean that for our major ongoing activities, the public knows when it can expect the data, and we deliver. Providing comparable data on a predictable basis not only increases their usefulness, but hopefully builds a greater understanding of their meaning and relevance through regular discussion. Key indicators should not come “out of the blue;” rather, the public should be waiting expectantly for them.

And by timely, I mean that activities are undertaken and data reported with an eye toward addressing key questions of current interest.

Finally, in addition to those three goals—usefulness, regularity, and timeliness—we strive to make our research as open as possible. Statistical research can be extremely complex, and not everyone is interested in the details on collection procedures and results. For those who are though, we try to ensure that—within

the bounds of protection of confidentiality—our methods and results are as open as possible. As you will soon hear me describe, in the case of TIMSS, I feel we have met this goal.

II. Background on TIMSS

NCES is but one of many organizations worldwide involved in the collaborative effort to develop, conduct, and report on the findings of TIMSS. The primary sponsor of the study is the International Association for the Evaluation of Educational Achievement, the “IEA,” which is headquartered in The Hague, in the Netherlands. Since its inception in 1959, the IEA has conducted studies to provide information and insight into the achievement and context of educational systems around the world. TIMSS is the first time that assessments of mathematics achievement and science achievement have been conducted as part of the same study. It represents the third study for each subject. TIMSS assessed students at three population levels: fourth grade, eighth grade, and at the end of high school. Altogether, the study involved roughly a half-million students from 41 countries, and the use of 30 languages, making it the largest and most comprehensive international study of education that has ever been undertaken.

To complement the IEA study of achievement, NCES and the National Science Foundation sponsored a study to analyze the mathematics and science curricula of the countries participating in TIMSS. NCES also sponsored a videotape study of teaching methods in eighth-grade mathematics in Japan, Germany, and the United States, and sponsored case studies of those same countries where teams of researchers spent several months at schools observing classes and speaking with students, teachers, administrators and parents. These additional studies, combined with the achievement studies and their surveys of students, teachers, and administrators, collectively provide a much fuller understanding of science and mathematics education around the world than we have ever had in the past.

Utilizing all of these studies related to TIMSS, NCES has written a series of books entitled *Pursuing Excellence*, with one volume for each of the three grade levels examined in TIMSS. The goal of the series was to look at all of the data gathered through the different components of TIMSS and synthesize and present the findings for an American audience.

Our purpose in publicizing the results of these studies is not to criticize or place blame, but to bring a different perspective to discussions of achievement, curriculum and instruction. That goal led NCES, in collaboration with the Department of Education’s Office of Educational Research and Improvement, to develop *Attaining Excellence*, a “tool kit” for use by state, district, and school staff. The Tool Kit contains summaries of the research, but also sample lessons from the videotape study and a guide to using the results of TIMSS to examine state, local, and school policy and practice.

The results of TIMSS have been released in several pieces, or perhaps more accurately, waves. Each has received a fair amount of publicity in the media, and you may recall different news stories, articles, and headlines from the past two years. The first wave came in the Fall of 1996 and contained the results of the eighth-grade achievement results and the curriculum study. The second wave came in June of 1997, with the release of the fourth-grade achievement results. The third wave came a few months ago in February, with the release of the achievement results for students at the end of high school. It is here I’d like to begin.

III. Design of End of Secondary School Assessments

The end of high school component sought to compare students in the final year of secondary education, as defined by the systems of each of the participating countries. Recognizing that education systems differ in terms of years of schooling and the age at which students typically complete secondary education, we still find such a comparison extremely valuable, as—regardless of age or years of schooling—the completion of

secondary school is often seen as the point at which young people are deemed “ready” to take their place in adult society.

The end of secondary school component included four assessments:

- 1) an assessment of general mathematics knowledge given to samples of students in 21 countries representative of the general student population,
- 2) an assessment of general science knowledge given to the same students,
- 3) an assessment of advanced mathematics knowledge given to samples of students in 16 countries representative of each country’s advanced mathematics students, and
- 4) an assessment of physics knowledge given to samples of students in 16 countries representative of each country’s physics students.

This study included primarily the industrialized countries of Europe but also the United States, Canada, and New Zealand. [**SHOW OVERHEAD 1—LIST OF COUNTRIES**] The slide shows a list of all countries participating in the general knowledge assessments. With a few exceptions, the sets of countries participating in the physics and advanced mathematics assessments came from this group.

The assessments were developed by an international committee of subject specialists. Assessment items were carefully reviewed by the participating countries to ensure that they reflected curriculum topics considered important in all countries, and did not over-emphasize the curriculum content taught in only a few. Although the two general knowledge assessments were not based on a particular subject-based curriculum, the content of the mathematics general knowledge assessment was estimated to include topics covered by the seventh grade in most countries, but not until the ninth grade in the United States. The content of the science general knowledge assessment was estimated to include topics covered by the ninth grade in most countries, but not until the eleventh grade in the typical U.S. curriculum. These assessments contained a mixture of question types, including multiple choice, short answer, and extended free response.

As was the case for the entire TIMSS study, the development, design, implementation, and analysis of the end of secondary school assessments were overseen by an International Steering Committee and several other international advisory committees, including a Technical Advisory Committee, a Subject Matter Advisory Committee, and a Quality Control Committee. In the United States, the Board on International Comparative Studies in Education of the National Research Council also monitored the various aspects of the study. These groups included not only the world’s leading experts on comparative studies of education systems, but also experts in assessment design and statistical analysis. Together, they helped ensure that TIMSS met the highest level of research standards. It is because of this attention to quality that I can say without hesitation that I am comfortable with the design and accuracy of the findings. Hopefully my faith in the quality of the study will be clear from my remarks, but I want to make sure I state this point up front.

How We Discuss Performance: Before I go further, I would like to speak briefly about how we talk about performance, that is, what we mean when we say U.S. students’ performance was “good,” “average,” or “poor.” First, at this point, we are only speaking in terms of U.S. students’ performance *relative* to students in other countries. Unlike letter grades on a report card or the Science Proficiency Levels we have used in NAEP, there are no definitions of what achievement of a particular score on the TIMSS assessment means. Thus we speak only of performance in relation to other countries participating. Doing so, however, is entirely consistent with our National Education Goal of being the “first in the world in mathematics and science.”

The other point I would like to make about looking at performance has to do with ranking countries. It is important to remember that because TIMSS is a sample study, the average scores of the populations tested are only *estimates* of what the score would have been had all the students in the country within the target

population been tested. And, as you, as scientists, are well aware, these types of estimations have margins of error. As a result, when one country's estimated score is higher than another's, or is higher than the international average, we cannot say for certain that this difference in scores or even the rank order would have been the same had all the students in the target populations been assessed. So, rather than rely on score differences and rankings, we establish levels of statistical significance and say a country is "higher" or "lower" than another, or than the international average, only if the difference is statistically significant.

So, as you will see in a moment and throughout my talk, we talk about U.S. performance in terms of the number of countries scoring *statistically significantly higher* than us, *statistically significantly below* us, and *not statistically significantly different* from us. We use similar definitions when discussing our performance relative to the international average. We believe comparisons that do not use these criteria—such as most "horse race" analyses of the United States's exact rank—are misleading, inaccurate, and should not be done.

IV. Response to Criticisms of the End of Secondary School Assessments

In my public statements, I have referred to our performance in the same way I just described them: "not good" and "among the lowest." Secretary of Education Richard Riley and President Clinton have both made similar comments. While many agree with us, a few people have criticized us for our interpretation of the results. This is no surprise. Our experience with international comparisons has given us an understanding of potential problems with comparing students across countries and the most common reactions to such comparisons. In fact, it is this experience that helped TIMSS researchers take steps to avoid such problems and conduct the study in such a way as to make it less vulnerable to these criticisms. TIMSS is not only the largest international study of mathematics and science achievement, it is also the most scientifically sound. Here I would like to briefly address a few of the most common criticisms.

Selectivity of the populations: Traditionally, the most common criticism of international comparative studies of achievement is that it is unfair to compare our high school students to students in other countries because a high school education in the United States is far more accessible to a larger percentage of the population. The argument goes that in other countries, toward the end of high school, only young people with high levels of academic achievement are still enrolled in school. Thus their national scores are based on a highly selective population. While this may have been true in the past, it is simply not valid in the case of TIMSS. Using several different methods of measuring enrollment, the various data available to us indicate that the enrollment rate in the United States is closer to the international average than to the upper extreme.

Furthermore, the theory that higher secondary enrollment rates hurt a country's overall achievement does not hold true in TIMSS. Looking at the results, we find that students in countries with higher enrollment rates than the U.S. tended to score significantly higher than the U.S. on both the mathematics and science general knowledge assessments. In TIMSS, the pattern generally appears to be that higher secondary enrollment rates are associated with higher levels of performance, rather than the reverse.

One problem in the past is that in some cases, countries that separate students by school according to academic ability and educational aspirations chose to test only those students in the more selective schools. Aware of the potential for this problem, TIMSS officials sought to ensure that the student samples in each country on the general knowledge assessments included students from all school types. In countries where students are separated into academic, vocational, and other types of tracks, students from all tracks were tested. Some countries, did, however, exclude certain students. The extent of this problem, however, has been overstated. Of the 21 countries participating in the general knowledge assessments, 14 excluded less than 10 percent of the internationally defined eligible population from the TIMSS sample. Furthermore, higher rates of exclusion did not appear to aid a country's performance. As

evidence, we see that five of the seven countries that excluded 10 percent or more of the eligible students performed the same or worse than the U.S. on one or both of the general knowledge assessments.

Age of Students and Years of Schooling: Some have claimed that it is unfair to compare U.S. high school seniors to students at the end of secondary school in other countries because those students are older and have more years of schooling. As I mentioned earlier, this component of TIMSS was less concerned about comparing students of comparable ages than it was about comparing young people at the point at which they are deemed ready to enter society. For those who are nevertheless concerned about the age difference, it was not as great as some have portrayed. Yes, the average age of our students was lower than the international average, but while some critics have focused on the only two countries where the average age of students tested was over twenty years, the international average age on the general knowledge assessments was much closer to the average age of our students: 18.7 years internationally compared to our students' average age of 18.1 years. The difference in ages was even smaller on the physics and advanced mathematics assessments: The average age of our students was 18.0 years for both assessments compared to the international average ages of 18.4 years in physics and 18.3 years in advanced mathematics.

It is also important to note that in some countries, this older average age of students reflects a later school starting age. In Denmark, Slovenia, Norway, Sweden, and parts of the Russian Federation and Switzerland, students start the first grade at the age of seven, compared to our typical starting age of six. Thus in some cases, students in other countries may have been older, but did not necessarily have more years of schooling.

Some students, however, did have more years of schooling. It is true that in other countries both college preparatory and vocational programs may go through grade thirteen or fourteen, but most of those countries also have other secondary programs that end earlier, often *before* grade twelve. While eight countries (out of 21 in the general knowledge assessments) included some students in grades above twelfth grade, all of those countries also included students in twelfth grade and five of them also included students below grade twelve.

The differences in average ages and years of instruction also need to be considered in light of the content of the general knowledge assessments. If the content were based on high-level curriculum topics, then younger students and students with fewer years of schooling might be at a disadvantage, since it is reasonable to think that they would be less likely to have been exposed to these topics than older students or students with more years of schooling. However, the TIMSS general knowledge assessments did not represent advanced-level content. Students in the United States and in other countries were being tested on topics they should have already covered, several years earlier in most cases.

To those who would still dismiss our performance as due to the difference in ages or years of schooling, I ask how they would then explain the fact that we were also outperformed by countries where the average age of students was *lower* than ours.

Cognitive skills tested: Another common criticism of this assessment is that it only tests low-level knowledge and thinking. In a few minutes, I will show you sample questions from the end of high school assessments. You can judge for yourselves whether the knowledge and cognitive skills required are indeed "low-level." At NCES we believe firmly that TIMSS is not a superficial test of knowledge and skills. Items on all four of the end of secondary school assessments included not only multiple choice questions, but also short and extended free-response items as well. Typical items on the mathematics assessment required students to analyze a stated problem, decide which mathematical tools to use to solve the problem, and then solve it. Science items presented students with an observation, situation, or hypothesis and asked them to use their knowledge of science to either explain the cause, predict the results, or describe how one might go about testing the claim. The items on the TIMSS assessments

required a strong base of knowledge, but also a variety of other intellectual skills, including reasoning, application of knowledge, and designing multi-step solutions.

Even if you disagree and feel that the TIMSS questions were low-level, either in terms of knowledge required or cognitive skill, I ask: Why should we breathe a sigh of relief that the assessments our students did so poorly on were “low-level?” What does this say about our students’ likely performance on tests of more sophisticated knowledge and skills?

I could go on at length about the validity of TIMSS, but the bottom line is this: TIMSS is the fairest and most accessible large-scale international comparison of educational achievement ever conducted. It was informed by the experiences and problems of previous studies, and in those areas, improves upon them. Of course, TIMSS is not a perfect study. For example, there continue to be problems with countries meeting established criteria for sampling especially in the end of secondary school assessments. Many of these problems, however, are not unique to TIMSS, as many large-scale sample studies often experience difficulty in meeting sampling criteria. For example in NAEP we have more difficulty meeting the sampling criteria at the twelfth grade than at the fourth or eighth grades. We recognize these problems, and in fact make every effort to indicate and explain these in our publications. And we at NCES feel these factors are simply not great enough to make the findings of TIMSS invalid.

TIMSS must be viewed as part of a continuum of the development of international comparative studies of education. TIMSS builds on previous work, and as studies are undertaken in the future, the experience of TIMSS will guide them and help lead to an even fuller understanding of education around the world. These efforts cannot be dominated by a single country or group of countries, but require the continued collaboration of the nations of the world. Conducting international assessments is a good example of an ongoing scientific inquiry, where each step builds on previous work and makes possible further exploration and understanding.

V. The End of Secondary School Achievement Results

The General Knowledge Assessments: As many of you already know from the media or from our published reports, U.S. students performed relatively poorly on all four end of high school assessments. This was not, however, a significant change from our performance on similar previous end of secondary school studies. I will summarize the results of all four assessments, but provide more detail on the science assessments.

In mathematics general knowledge [*SHOW OVERHEAD #2—MATH. GK ABOVE/BELOW*], as the overhead illustrates, of the 20 other countries participating in the assessment, the students in fourteen countries scored significantly higher than ours; the scores of students in four countries were not significantly different from ours; and students in only two countries scored significantly lower than our students. Our students’ average score was 461, below the international average of 500, a difference that is statistically significant. It is these comparisons that cause us to say we are “near the bottom.” Countries at the top included the Netherlands, Sweden, Denmark, and Switzerland.

We did better in science general knowledge. [*SHOW OVERHEAD #3—SCIENCE GK ABOVE/BELOW*]. As you can see from the overhead, fewer nations outperformed us (eleven in science as opposed to fourteen in mathematics), more were similar to us (seven as opposed to four), but we still only outperformed two countries. Our average score of 480 was closer to the international average, but still significantly below it. Here, we perform most similar to a band of countries that includes several of our economic competitors, notably France and Germany. Countries at the top included Sweden, the Netherlands, Iceland, and Norway.

[SHOW OVERHEAD #4, PERCENTILE CHART, SCIENCE GK] This chart provides a more detailed look at the distribution of our scores and of some other countries. For each country, the whole bar shows the range of scores between the 5th and the 95th percentiles. Markers within each bar show the 25th percentile level, a confidence interval around the mean, and the 75th percentile. As you can see, compared to most of the other countries, the bar for the U.S. shows a general shift downward for all markers.

The length of the bars gives us an idea of the range of students' scores. Many feel we have a much greater diversity of achievement in this country. As you can see on the chart, the range of scores in the United States from the 5th percentile to the 95th does not appear to be particularly large. While France and Canada have smaller ranges, we are comparable to most of the others.

As I'm sure you're interested in the type of questions on this assessment, I have some sample questions. For those of you who are interested in seeing more actual items, each volume of *Pursuing Excellence* contains sample items. **[SHOW OVERHEAD #5 WEB SITE]** The examples I am about to present are not in *Pursuing Excellence*, but can be found in the much larger sets of released items at <http://www.csteep.bc.edu/TIMSS1/Items.html>. Item sets are available in PDF format for all TIMSS assessments at all three grade levels.

This first example **[SHOW OVERHEAD #6—G.K. SAMPLE 1]** asks students to identify ways in which a student could have caught influenza. Students were given credit if their answer referred explicitly or implicitly to the transmission of germs, or if they stated only that he could have caught it from someone else with the flu.

As you can see, the majority of our students (59 percent) answered this question correctly, as did most students in the other countries. Relatively few of our students responded with the most common incorrect response internationally, that he caught it from being too cold.

This next question **[SHOW OVERHEAD #7—G.K. SAMPLE 2]** was slightly more difficult for students internationally, but still, a majority of our students answered it correctly. It asks students to explain why a stone thrown at a window would crack it, while a tennis ball thrown at a similar window would not. Students were given credit if their answer referred to the longer impact time and therefore smaller force of the tennis ball, for indicating that some of the kinetic energy of the ball was used to compress the ball, rather than all of it being used to break the glass, or for indirect references to those two concepts, such as the softness or deformation of the tennis ball or to the smaller impact area of the stone.

Similar to patterns in other countries, about 18 percent of our students provided an incorrect response that referred only to the differences in mass or density, or to the "sharpness" of the stone, but did not connect these ideas to impact time or the differences in kinetic energy.

On this last item **[SHOW OVERHEAD #8 GK. SAMPLE 3]**, students in most countries had quite a bit of difficulty.

Students were asked about the amount of light energy produced by a lamp in relation to the amount of electrical energy used. Although the question asked for an explanation, students were given credit for just choosing the correct answer, even without providing an explanation. Among U.S. students, only 11 percent received credit, compared to 21 percent internationally. Among our students receiving credit, only about one-fourth could provide a correct explanation.

The Advanced Assessments: The advanced assessments sought to compare students with advanced coursework on content at a level higher than "general knowledge." Each country used their own definition of "advanced," but were given the general guideline that the identified population of students should represent between 10 and 20 percent of the appropriate age cohort. In the United States, advanced mathematics students were defined as students who had taken or were taking a full year of a course that

included the word “calculus” in the title. These courses included pre-calculus, calculus, Advanced Placement calculus, and calculus and analytic geometry. The resulting population represented about 14 percent of the age cohort. For all countries, the advanced mathematics students represented about 19 percent of the age cohort.

Looking at our scores on the assessments for advanced students [*SHOW OVERHEAD #9—ADVANCED MATHEMATICS ABOVE/BELOW*], of the fifteen other countries participating in the advanced mathematics study, students in eleven of those countries scored significantly higher than ours, students in four countries scored similarly to ours, and students in none of the countries performed significantly lower than ours. Our average score of 442 was significantly lower than the international average of 501. In this case, our performance is not better than any of the participating countries.

Because Advanced Placement courses provide us with a commonly understood, more selective definition of “advanced”, we thought it would be useful to see how our students in AP Calculus measured up to other countries’ most advanced students. Keep in mind, though, that we are now making our group of elite students more selective, but are also making no corresponding adjustments to the populations in other countries. Approximately five percent of the U.S. age cohort takes AP Calculus, as compared to an average of 19 percent who were identified as advanced mathematics students in all the countries.

[*SHOW CHART #10—AP CALCULUS ABOVE/BELOW*] This analysis provides us with a different view: Our AP calculus students are competitive with the top students of the other participating countries. Our AP students’ score was 513, compared to the international average of 505. Furthermore, students in only one country, France, performed significantly higher than ours.

[*SHOW CHART #11—ADVANCED MATH CONTENT AREAS*] We were also able to take advantage of an extremely useful feature of TIMSS: the ability to disaggregate data by content area. Doing so in advanced mathematics, we see that despite the fact that our population of advanced mathematics students contained some students who had not taken calculus, our weakest content area was not calculus, but geometry. This is a rather troubling surprise, as our typical advanced mathematics student should have already taken a full year of geometry well before their final year of high school.

The assessment for advanced science students was a physics assessment. Accordingly, the population the U.S. identified for this assessment consisted of high school seniors who had taken or were taking at least one year-long course in physics. Course titles included physics I, physics II, advanced physics, and Advanced Placement Physics. This population also represented approximately 14 percent of the age cohort, which was the same as the international average.

Of all four end of secondary school assessments, our performance on the physics assessment was perhaps most disappointing. [*SHOW OVERHEAD #12—PHYSICS ABOVE/BELOW*] Here, students in fourteen of the fifteen other participating countries performed significantly higher than our students, with the only remaining country—Austria—scoring similar to the U.S.. There were no countries where students scored significantly below ours. Our students’ average score of 423 was significantly below the international average of 501. Just as in advanced mathematics, we were “at the bottom,” only this time, we had less company.

We also thought it would be useful to compare our most advanced students in physics, meaning those taking Advanced Placement Physics, to the advanced science students from other countries. Just as in advanced mathematics, our population of AP students is on average smaller than the populations of advanced science students in other countries, only in the case of physics the difference was much greater: whereas advanced science students represented an average of 14 percent of the age cohort in other countries, only 1 percent of the U.S. age cohort take AP Physics.

[SHOW CHART #13—AP PHYSICS ABOVE/BELOW] As you can see from the chart, even comparing only the “best of our best” to other countries’ “best”—“stacking the deck,” if you will—we still fail to reach above the international average. Rather than outperforming no countries, as our population of all physics students did, our A.P. Physics students do only slightly better by outperforming one country, Austria. We are now in a larger band of similar scoring countries, but there is still a substantial number of countries outperforming us. Our most advanced 1 percent is outperformed by Sweden’s most advanced 16 percent and Australia’s most advanced 13 percent. This finding should wake up anyone who still believes our most advanced physics students are competitive with the most advanced students in other countries.

[SHOW CHART #14—PHYSICS CONTENT AREAS] Looking at the content areas of the TIMSS physics assessment, we see that while we performed poorly in all areas, Mechanics, and Electricity and Magnetism were our weakest areas.

[SHOW OVERHEAD #15—PERCENTILE CHART FOR PHYSICS] Returning to overall performance on the physics assessment, this chart of the distribution of scores shows that, similar to science general knowledge, the performance of our physics students was lower at all percentile levels than most of the other countries shown.

Looking at the range of scores in the United States (from the 5th percentile to the 95th percentile), we see that it is much smaller than the other countries on the chart, as was the case for most of the countries participating in the assessment.

I also have some sample items for the physics assessment. This first one **[SHOW OVERHEAD #16—PHYSICS SAMPLE ITEM 1]** is an example of a Heat item. It asks students to choose an explanation for the increase in volume as water boils to produce steam. The majority of our students, 60 percent, chose the correct answer, compared to an international average of 65 percent. Roughly one-quarter of our students mistakenly chose “B,” which stated that water molecules expand when heated. This answer was much less common internationally than it was in the United States.

This next item **[SHOW OVERHEAD #17—PHYSICS SAMPLE ITEM 2]** is an example of an Electricity and Magnetism item, one of our weakest content areas. It was difficult for most students, as only 32 percent of students in all countries tested answered it correctly, but it was even more difficult for our students, of whom only 12 percent received credit. The question asks students to choose the correct path of electrons entering an electric field. The most common answer internationally and among our students as well, was choice “B,” path II. The next most popular answer among our students was choice “C,” path III.

This last example **[SHOW OVERHEAD #18—PHYSICS SAMPLE ITEM 3]** is a Mechanics item, another one of our weakest content areas. It was extremely difficult for all students. Internationally, only sixteen percent received credit. Among U.S. students, only six percent received credit. It asked students to draw arrows on the figure showing the direction of the acceleration of the three points indicated. Over half of our students either drew arrows along the trajectory of the ball, or drew two arrows for each point, one pointing either up or down, and the other perpendicular to it, pointing forward.

VI. What TIMSS Has Shown Us – Achievement

Now that I have presented the most recent findings and addressed some of the criticisms of the study you may have heard, I would like to take a broader look at what all the components of TIMSS have collectively told us about science achievement and education in the United States. I will begin with observations about achievement and then examine whether TIMSS has shed any light on the causes for our performance.

So, looking at the different results of TIMSS that have been released over the past two years, what is the “story” regarding science achievement in the United States? I believe there are three major stories to tell:

One is that despite generally positive signs at the fourth grade level, by the time our students are ready to leave high school—ready to enter higher education and the labor force—they are doing so with an understanding of science that is significantly weaker than their peers in other countries.

The second story is that our idea of “advanced” is clearly below international standards.

And the third story is that there appears to be a consistent weakness in our students’ performance in physical sciences that becomes magnified over the years.

In addition to these three main points, I would like to address the question of gender gaps in achievement, which appear to be less of a problem in the United States than in other countries.

Achievement over the years: I have already presented you with the results of the end of high school assessment. Just to refresh your memory, they were not good. However, the results from the fourth- and eighth-grade assessments were—at least at the time of their release—somewhat more encouraging.

On the fourth-grade assessment [*SHOW OVERHEAD #19—4TH GRADE ABOVE/BELOW*], our students performed very well. Our students’ average score was significantly above the international average, and of the students in the 25 other participating countries, students in only one country, Korea, outperformed ours. Our students outperformed students in 19 other countries, including Canada, Singapore, and Norway, and there was no significant difference between our students’ scores and the scores of students in five other countries, including Japan. By the same criteria used to say we were “near the bottom” at the high school level, we appear to be “near the top” in elementary school.

This picture changes somewhat as we move on to middle school. [*SHOW OVERHEAD #20—8TH GRADE ABOVE/BELOW*] On the eighth-grade science assessment, there was a larger number of countries participating: 40 countries in addition to the United States, compared to a total of 26 countries in the fourth-grade assessment. On this assessment, our students still scored significantly above the international average, but our students were outperformed by nine countries. Sixteen other countries scored similar to us, and 15 countries scored significantly below us. In this case, it is far more difficult to say we are “near the top.”

Now, to complete the picture, let’s look again at the twelfth-grade general knowledge assessment. [*SHOW OVERHEAD #21—12TH GRADE SCIENCE G.K. ABOVE/BELOW*] Of the 20 other countries in the study, we were outperformed by a majority, 11, and outperformed only two. The students in seven countries scored similar to us. And our students scored significantly below the international average.

So that I cannot be accused of trying to mislead you, let me point out that these were three different assessments, involving three different sets of countries. Only thirteen countries participated in the assessments at all three grade levels. Thus, one could argue that there may in fact be no trend of poorer relative performance at the higher grade levels, and that if all countries had participated in all of the assessments, the results might have been quite different. I will not argue that point. Until we have such a perfect study, though, we will continue to use the data on the different sets of countries that participated at each level to define our international performance goals.

Our idea of advanced: For years, people have taken comfort in the notion that while the performance of all our students may be only average, our strength lies in our top students. Many people feel that our best students perform better than the best students of most other countries. I believe TIMSS shows this notion to be untrue. Again, the findings of the early years are somewhat encouraging.

[SHOW OVERHEAD #22—PERCENT OF STUDENTS THAT WOULD BE IN TOP 10%] One way to establish a benchmark for “top students” is to identify the 90th percentile for all the students taking the test. This means that 10 percent of the international pool scored above this mark. If we in fact have a disproportionately large share of elite students, we should have more than 10 percent of our students reaching that mark. At both the fourth- and the eighth-grade levels, we do in science. At the fourth-grade level, 16 percent of our students would qualify, and at the eighth-grade level, 13 percent of our students would qualify.

While similar data for the twelfth-grade general knowledge assessment have not been calculated, we can look at the question in another way by referring back to our distribution chart. **[SHOW OVERHEAD #23—SCIENCE G.K. PERCENTILE DISTRIBUTION]** If it were true that our top students perform at higher levels than top students in other countries, then we would expect to see these lines here (indicating our 75th and 95th percentile scores) higher than in the other countries. They are not. In fact, they are generally below them.

Here I would like to point out again that even if we use the highly selective group of our A.P. Physics to compare with physics students from other countries, we still perform poorly. **[SHOW OVERHEAD #24—ABOVE/BELOW CHART FOR AP PHYSICS COMPARISON]** The scores of our top one percent of the age group are no different from the scores of students who represented far less selective groups of students in other countries. Their scores are significantly higher than students in only one other country.

What does this tell us? To me, it says that our concept of “advanced” is different—and lower—than in other countries. Furthermore, I believe that this is a gap that widens over time. We have empirical evidence from the increasingly lower relative international standing of our students from the fourth to the twelfth grade in both mathematics and science. And we have further support for the idea of a widening gap from the curriculum experts who tell us that topics typically covered by ninth grade in most countries are not covered until the eleventh grade in the United States. If these assessments are correct, then we may be suffering from a curriculum drift, where the level of rigor of our curriculum fails to keep pace with international standards during the years of middle school and high school. The impact of this possible gap on achievement would be obvious. I ask you, the science education community, to examine science curricula and achievement to determine whether we are maintaining an appropriate level of rigor across grade levels.

Consistent weakness in physical sciences: Our poor performance on both the physics assessment and the twelfth grade general knowledge assessments came as a surprise to many, in view of our “near the top” performance in fourth grade and “above average” performance in eighth grade. If we look hard at those results however, and break them down by content area, it is possible to see a foreshadowing of our poor performance at the end of high school.

Both the TIMSS fourth and eighth grade assessments can be broken down into content areas. In elementary school, there were four content areas: Earth Science, Life Science, Physical Science, and Environmental Issues and the Nature of Science. **[SHOW OVERHEAD #25—4TH GRADE CONTENT AREAS ABOVE/BELOW]** In all four content areas, our students scored above the international average. However, if we were to identify our weakest content area, it would be physical science. My definition of weak has to do with the number of countries whose students scored significantly higher, significantly lower, and not significantly different from ours.

When we look at eighth grade science content areas **[SHOW OVERHEAD # 26—8TH GRADE CONTENT AREAS ABOVE/BELOW]** again, we see our weakest performance was in the physical sciences, which in this case were represented by two content areas, chemistry and physics. Compared to our performance in the other content areas, in physics and chemistry we were outperformed by more countries and outperformed fewer countries. Furthermore, we were not above the international averages.

So, if the physical sciences are our weakest areas in both fourth and eighth grades, should it come as a surprise that we do so poorly in physics in twelfth grade? Unfortunately, the twelfth grade science general knowledge assessment was not organized by content area. So, perhaps I can only offer this as a theory rather than an observation. Hopefully this question can be explored in greater detail either in secondary analysis of TIMSS data, or in future studies. Since you are the experts, I call upon you to examine science curricula and instruction in the United States to see if there might be reason to suspect the consistent weakness in the physical sciences shown by TIMSS.

Gender gaps in some areas: In addition to the three major “stories,” I would like to tell you what we found regarding differences in scores between boys and girls. Looking at all of the results, even though gender gaps existed at some of the grade levels and in some of the content areas, I think the results are, on the whole, encouraging.

On the fourth-grade science assessment, U.S. boys’ overall scores were significantly higher than girls. This gap in overall scores appears to be driven by the significant gaps demonstrated in the content areas of earth science and physical science. Of the 25 other countries participating in the study, nine countries in addition to the United States had gender gaps in overall science achievement.

At the eighth-grade level we find our first piece of encouraging news: There were no significant differences in scores in the United States between boys and girls—not on the overall scores, and not in any of the categories. We were one of eleven countries where no gender gaps existed, of a total of 41 countries participating.

At the twelfth-grade level, on the general knowledge assessment, of the 21 participating countries, South Africa was the only country where there was no significant difference between scores of males and scores of females. Similarly, on the physics assessments, males scored significantly higher than females in the United States, and all other countries except Latvia. These gaps are of course not acceptable, but we should take some encouragement in the fact that our gender gaps at the twelfth-grade level are among the smallest of all countries on both the general knowledge and the physics assessments. Additionally, while the populations are not perfectly comparable, looking at earlier IEA assessments of science achievement, we believe we are seeing gradual improvement in this area.

VII. Explanations for Achievement From TIMSS Data

It is perhaps only natural that the achievement data—“the results”—are what grab people’s attention. Ranks and scores make good headlines, and are as easy to comprehend as results of an election, or a horse race. But it is my sincere hope that people’s attention will stay focused long enough to explore the more difficult question of *why* our students performed as they did. The better we can understand this, the better able we are to help our students.

We can begin to look for explanations in TIMSS itself. One way in which TIMSS improves upon previous comparative studies is the extent to which it looked beyond achievement to gather data on the context of education. Through detailed questionnaires given to students taking the assessments and their teachers, TIMSS gathered data on science instruction and on students’ habits in and out of school. And through the TIMSS curriculum study, we now have detailed information on the content of textbooks and curriculum frameworks found in the TIMSS countries.

I will tell you now, though, that if you are looking for “the reason” why our students performed as they did, you will be disappointed. What TIMSS does offer are the data to begin developing hypotheses regarding our patterns of achievement. I will present some of these hypotheses, but I encourage this audience to become active users of the TIMSS data to confirm previously presented statements and to develop new ideas.

Curriculum: A logical place to start looking for explanations for our performance is in our curriculum. Here, while it may not necessarily provide the answer, the TIMSS curriculum analysis serves as an excellent jumping off point for further exploration of our science curriculum.

The curriculum study analyzed textbooks and state and national curriculum frameworks. The analysis focused on the topics in the curriculum, such as “Energy Types, Sources, and Conversions,” or, “Biomes and Ecosystems.” It looked at not only the number of topics, but the topic contents as well. This was the first time a study of this type and of this size had been conducted. Thus far, the findings regarding topics and their relationship to achievement are perhaps a bit more clear in mathematics than in science. In mathematics, researchers found that at all levels, the typical U.S. curriculum contained a far greater number of topics than most other countries. They saw a strong connection between our broad curriculum and our poor performance in mathematics.

In science, while the findings were similar to mathematics, their connections to achievement are at this point less clear. Similar to mathematics, the study found that our science textbooks contained a much larger number of topics than the international average for the three grade levels assessed by TIMSS. A comparison of state curriculum frameworks in the U.S. with state and national curriculum frameworks from other countries found the typical U.S. curriculum to have more topics than most countries through the beginning of high school, after which the number of topics in our states’ curriculum frameworks begins to fall below the international average. Where we must be careful is in considering the various types of science courses that exist across the United States, particularly at the middle school level. I’m sure among this audience we have middle school teachers who teach across the science domains during a year-long course, but also ones that concentrate on physical science, life science, or earth science. Thus, especially at the middle school level, the large number of topics found in the textbooks and curriculum guides may be more a result of the wide range of offerings in science found in this country than of the number of topics faced by the typical student.

Still, the curriculum study has created a very healthy discussion of our science curriculum, focused on the simple yet fundamental question of “Do we have too many topics in our curriculum?” This is an example of the power of an international study: Comparing ourselves to other countries can help us question and examine our traditional ways of doing things.

Course-taking and hours of instruction: Another logical place to look for answers would be to ask whether our students have the same amount of science instruction as their peers in other countries. In some cases they do, and in some cases they do not. For example, our fourth-grade students receive on average 48 more minutes of science instruction per week than the international average. Our students receive an average of 2.7 hours per week, compared to the international average of 1.9 hours a week. Students in only one country, Portugal, received more average hours of instruction per week. These data are based on teacher surveys, and only look at cases where science is taught as a separate subject, which it is in most classes in most countries. We do not have comparable data at the eighth grade level, but other studies indicate that our eighth grade students have more hours of science instruction per year than their counterparts in either Germany or Japan.

At the twelfth-grade level, surveys of students taking the general knowledge assessment did not ask about hours of instruction, but simply whether they were taking a science class that year. On average, fewer of our students were taking science as compared to their international peers. Forty-seven percent of our twelfth graders reported that they were not taking a science class that year. This was significantly higher than the international average of 33 percent of students in the final year of secondary school.

On the surface, these numbers would appear to help explain our performance: our students have more hours of science instruction in the fourth grade and they perform well, whereas a relatively large proportion of our students are not taking science in their senior year, and they perform poorly. However,

in looking at the twelfth-grade data, we see that some of the top performing countries actually had *larger* proportions of their students *not* taking science. For example, in Sweden, which had the highest average student scores on the general knowledge assessment, 57 percent of the students at the end of secondary school were not taking science, compared to 47 percent in the United States. In fact, when we ran a statistical analysis of the data, we found that our relatively large number of students not taking science was not a factor in our relatively poor performance.

Hours of Homework: If it is not hours of instruction in school, perhaps our students' performance is connected to how much time they are devoting to science *outside* the classroom. Here again, though, there appears to be very little connection. In fourth grade, where our students' performance was near the top, on a normal school day, our students reported spending an average of 48 minutes outside of school studying science or doing science homework. This was roughly equal to the average of all other countries. At the end of high school, where our students' performance was near the bottom, among those students taking science, the data on homework were similar: our students appeared to be spending no more or no less time than their international peers on science homework. I should note that on the twelfth-grade survey, the questions did not specify out-of-school time, thus creating the possibility that our students included time they spent in class working on homework, a practice some have observed is considerably less common in other countries.

One interesting note—and perhaps disturbing to some—is that the amount of science studying and science homework for students taking science in the twelfth grade is less than one hour on a normal school day, both in the United States and in the other countries.

Instructional strategies and lesson content: Although limited, the TIMSS surveys provide a glimpse into the type of instruction our students receive. While I am not going to try to provide you, the National Science Teachers Association, with a model for the best way to teach science, I feel it is safe to say that good science instruction is, among other things, inquiry-based, prods students to think, and makes connections to the world in which we live. The TIMSS survey of physics students found that, at least from the student point of view, those qualities fit physics instruction in the United States. U.S. students appear more likely to describe their science instruction in these terms than physics students in other countries. Unfortunately, this cannot help keep them out of the international achievement basement.

Attitudes toward science: We find similarly positive response patterns when we look at surveys of students' attitudes toward science. Looking across all the student surveys in TIMSS, generally speaking, our students say they like science, believe it's important for their futures, believe their parents think it's important, and—even at the high school level—believe they are doing well in science.

Summary: Taking into account the data from student and teacher surveys and from the curriculum survey, our major finding regarding the cause for our patterns of achievement—whether it be our students' relatively strong performance in fourth grade or their relatively weak performance in twelfth grade—is that there is no one single, readily apparent explanation. Student achievement reflects a complex array of influences, some of which were examined by TIMSS, but many of which were not. The role these influences play, and how they interact with each other, will be a primary focus for future analysis of TIMSS data.

VIII. Questions Raised By TIMSS

The results from TIMSS raise several provocative questions about our science curriculum, textbooks, and instruction. Here, I will offer several questions, focusing on those that pertain specifically to the schooling process. I hope these questions can serve as starting points for further exploration not only within my

agency, but among other organizations, the research community, policymakers, and among you, the educators.

How rigorous is our science curriculum? Compared to their international peers, our students seem to be getting as much instruction, are studying as much and are doing as much homework. On top of it all, they claim to like science and recognize its importance. Then why is it that by the end of high school, our students are being outperformed by students in so many other countries, some of whom are younger, studying less, or are not as likely to be taking science at the end of secondary school. If our students are doing so many things right, or at least no different from their international peers, isn't it possible that we are feeding them a relatively weak science diet? Our curriculum analysis just scratched the surface of this issue. I hope the TIMSS findings will result in a more intense examination of whether what we are expecting our students to learn is really on par with the rest of the countries of the world.

Is our “stop-and-go” curriculum hurting us? An issue closely related to the level of rigor of our science curriculum is the structure of our science courses. Starting as early as the seventh grade in many places, students' study of science is specialized: one year they may study earth science, the next year, biology perhaps, and the following year, maybe chemistry. This is *not* the way it is done in many countries. A more common model is for students to study more than one science area simultaneously. The end result is that students in the United States may focus on physical sciences, for example, during a given year, but one or more years may go by before they pick up that topic again, that is, if they are still taking science. Contrast this with the more consistent approach in other countries, where students face a steady curricular stream that allows topics to build on one another, year after year.

Again, our role at NCES is to provide information that can help inform key policy decisions. Looking at other countries shows us alternate approaches to the structure of science curriculum. Recognizing that this is an important issue within the science education community, I hope findings from TIMSS can be useful and that future research efforts explore this area further.

Are our percentages of teachers teaching out-of-field higher than in other countries?

Data from another NCES study shows us that a significant number of students are enrolled in science classes taught by someone who has neither a college major nor a minor in the subject. If we look at all high school science teachers, we find that approximately eighteen percent of them do not have at least a minor in either science education or any of the science specialties. This percentage was actually among the lowest of the subjects studied. However, when we look more closely at background in specialty areas, the numbers are quite different. Among teachers of high school biology and life sciences classes, approximately 31 percent of them do not have at least a minor in biology. Among high school physical science teachers, over half, 55 percent, do not have at least a minor in any of the physical sciences. Unfortunately, this is an area where we do not yet have good international data. But because these numbers appear high, it would be interesting to see how they compare to similar data in other countries.

We have some indications that these numbers are not as high in other countries. One interesting story that was told to me was that when the TIMSS questionnaire design groups met to develop the teacher questionnaire, panel members from the United States suggested including an item regarding whether the teacher had an educational background in the field they were teaching. This suggestion was greeted with laughter, as the other members could not understand why such an item would be necessary. It was not included.

Are our teachers receiving adequate pre-service education and appropriate support in their first few years of teaching?

From TIMSS and other research, we know that pre-service and teacher induction practices vary from country to country. While our teachers are among the most educated—in terms of postsecondary degrees—others countries sometimes require more classroom experience prior to certification or provide more support for new teachers. For example, in Germany, teachers must serve what amounts to a two-year teaching “apprenticeship” before they are granted

certification. And in Japan, teachers in their first full year of teaching have a reduced class load to accommodate professional development activities, both in school under the guidance of a mentor teacher, and at a centralized professional development facility. Again, we do not have yet have a complete picture, but hope that the initial research can be of relevance to those exploring this issue in the United States.

IX. Future TIMSS Activities

In my talk today, I have tried to summarize the major findings of TIMSS related to science education and achievement in the United States. In TIMSS, we have the richest source of information ever collected on science and mathematics education around the world. It has already taken us to a new level of understanding of our own education system and others, and has raised many important questions about educational practice. At NCES, we have been working furiously to present the results of TIMSS in an accessible format as soon as they are available to us. I am proud to say that within a year of the release of the first wave of TIMSS data—the eighth-grade assessment results—my agency was able to turn the data into a series of products to help people better understand and utilize the findings of TIMSS. The TIMSS Resource Kit is just one example. Our hope is to continue on this path of exploration to bring us to an even higher level of awareness and perhaps to some policy recommendations. I would like to leave you with just a few of the future activities related to TIMSS.

Secondary analysis: We have only just begun to tap the full potential of the TIMSS data. We hope that further analysis of TIMSS can be an undertaking shared by all. To the extent possible, we are making the procedures and data of TIMSS open and accessible. I invite you and your colleagues to the TIMSS “gateway” World Wide Web site at <http://nces.ed.gov/timss>. This site contains links to many TIMSS resources including a large number of actual assessment items. The millions and millions of bytes of data are also being put into a database for use by researchers. NCES is encouraging the further analysis of TIMSS data by funding training sessions for researchers on what is contained in the database and how they might use it. As we speak, researchers in colleges and universities throughout the country and the world are already busy formulating and testing a whole host of policy-relevant hypotheses, in areas such as the connection between achievement and student characteristics and habits, and differences in achievement on different types of questions.

Grade 8 NAEP/TIMSS Link: One way in which we plan to use the TIMSS data is to link it to the results of NAEP, the National Assessment of Educational Progress. Although the questions were different, we have developed a model for predicting success on one assessment on the basis of the actual results of the other. In May, we will release comparisons for grade eight mathematics and science. Using these comparisons, individual states will be able to see which countries their performance is most comparable to, and which countries perform at a level above or below them.

Second round for 8th graders in 1999 (TIMSS Replication): Not only will there be more analyses of the TIMSS data, but studies similar to TIMSS are already scheduled for the near future. In 1999, a replication of the TIMSS assessments will be given to eighth graders in the U.S. and in other countries. While these are not necessarily the same students, this is the same age cohort that participated in the TIMSS fourth-grade assessment in 1995. We also expect many of the same countries to participate on this new assessment as participated on the original TIMSS assessment. On this assessment, we plan to give states and districts the option to give the assessment to their own students at the same time as the international study, so that they might better understand whether their students are truly world class.

Videotape study of science instruction: As I mentioned earlier, NCES sponsored a videotape study of eighth-grade mathematics instruction in Japan, Germany, and the United States. Thus far it has proven to be an excellent source of information for both teachers and researchers. Teachers have had an opportunity to see typical instructional methods in other countries, and researchers have been able to

analyze the lessons to formulate hypotheses about how teaching in the United States differs from that in other countries. One major finding was that U.S. mathematics teachers tended to state key concepts, whereas teachers in Japan tried to develop the concepts, in part through an inquiry-based approach. We know that science instruction in the United States is very different from mathematics instruction, thus we hesitate to draw conclusions about it based on this study. Therefore, it is my goal to see a similar videotape study of science instruction. I believe that just as the mathematics study pointed to key differences in teaching methods in mathematics, a study of science instruction, hopefully in more and different countries, will be equally enlightening. It may answer such questions as, "How do teachers ensure that inquiry-based, hands-on instruction is balanced with a firm understanding of fundamental scientific facts and theories?" Or, "How do teachers integrate the different content areas within science, draw connections to material previously learned, and help students see the connections to the world around them."

X. Conclusion

In conclusion, I hope that I have conveyed to you my excitement about TIMSS, my concern for what it has told us about science education in the United States, and my hopes for future discussions and study. I know that as science teachers, you share these feelings as well. We live in times when our world is being transformed by revolutionary developments in science and technology. You are the ones with the extremely crucial, yet immensely challenging task of preparing our students not only to take their places as the scientific leaders in that world, but also to live in it. Many have argued that science is not vital to a person's education, that they can survive just fine in the world without a high level of scientific knowledge. I am sure you have heard these arguments yourselves, probably from a fair share of your students. I am not here to preach to you, the choir, that science education is important. We at NCES have already demonstrated our belief in the importance of science education through our commitment to TIMSS. In the future as well, we hope to continue to provide you with the kind of information that can help you fulfill your mission.

FINAL YEAR OF SECONDARY SCHOOL

21 PARTICIPATING NATIONS IN THE TIMSS GENERAL KNOWLEDGE ASSESSMENTS:

AUSTRALIA

ICELAND

SWEDEN

AUSTRIA

ITALY

SWITZERLAND

CANADA

LITHUANIA

UNITED STATES

CYPRUS

NETHERLANDS

CZECH REPUBLIC

NEW ZEALAND

DENMARK

NORWAY

FRANCE

RUSSIAN FEDERATION

GERMANY

SLOVENIA

HUNGARY

SOUTH AFRICA

FINAL YEAR OF SECONDARY SCHOOL

MATHEMATICS GENERAL KNOWLEDGE ACHIEVEMENT

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY HIGHER THAN THE U.S.	
NATION	AVERAGE
(NETHERLANDS)	560
SWEDEN	552
(DENMARK)	547
SWITZERLAND	540
(ICELAND)	534
(NORWAY)	528
(FRANCE)	523
NEW ZEALAND	522
(AUSTRALIA)	522
(CANADA)	519
(AUSTRIA)	518
(SLOVENIA)	512
(GERMANY)	495
HUNGARY	483

NATIONS WITH AVERAGE SCORES NOT SIGNIFICANTLY DIFFERENT FROM THE U.S.	
NATION	AVERAGE
(ITALY)	476
(RUSSIAN FEDERATION)	471
(LITHUANIA)	469
CZECH REPUBLIC	466
(UNITED STATES)	461

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY LOWER THAN THE U.S.	
NATION	AVERAGE
(CYPRUS)	446
(SOUTH AFRICA)	356

INTERNATIONAL AVERAGE=500

NOTE:

Nations not meeting international sampling and other guidelines are shown in parentheses.

SOURCE: Mullis et al. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School*.
Table 2.1. Chestnut Hill, MA: Boston College.

FINAL YEAR OF SECONDARY SCHOOL

SCIENCE GENERAL KNOWLEDGE ACHIEVEMENT

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY HIGHER THAN THE U.S.	
NATION	AVERAGE
SWEDEN	559
(NETHERLANDS)	558
(ICELAND)	549
(NORWAY)	544
(CANADA)	532
NEW ZEALAND	529
(AUSTRALIA)	527
SWITZERLAND	523
(AUSTRIA)	520
(SLOVENIA)	517
(DENMARK)	509

INTERNATIONAL AVERAGE=500

NATIONS WITH AVERAGE SCORES NOT SIGNIFICANTLY DIFFERENT FROM THE U.S.	
NATION	AVERAGE
(GERMANY)	497
(FRANCE)	487
CZECH REPUBLIC	487
(RUSSIAN FEDERATION)	481
(UNITED STATES)	480
(ITALY)	475
HUNGARY	471
(LITHUANIA)	461

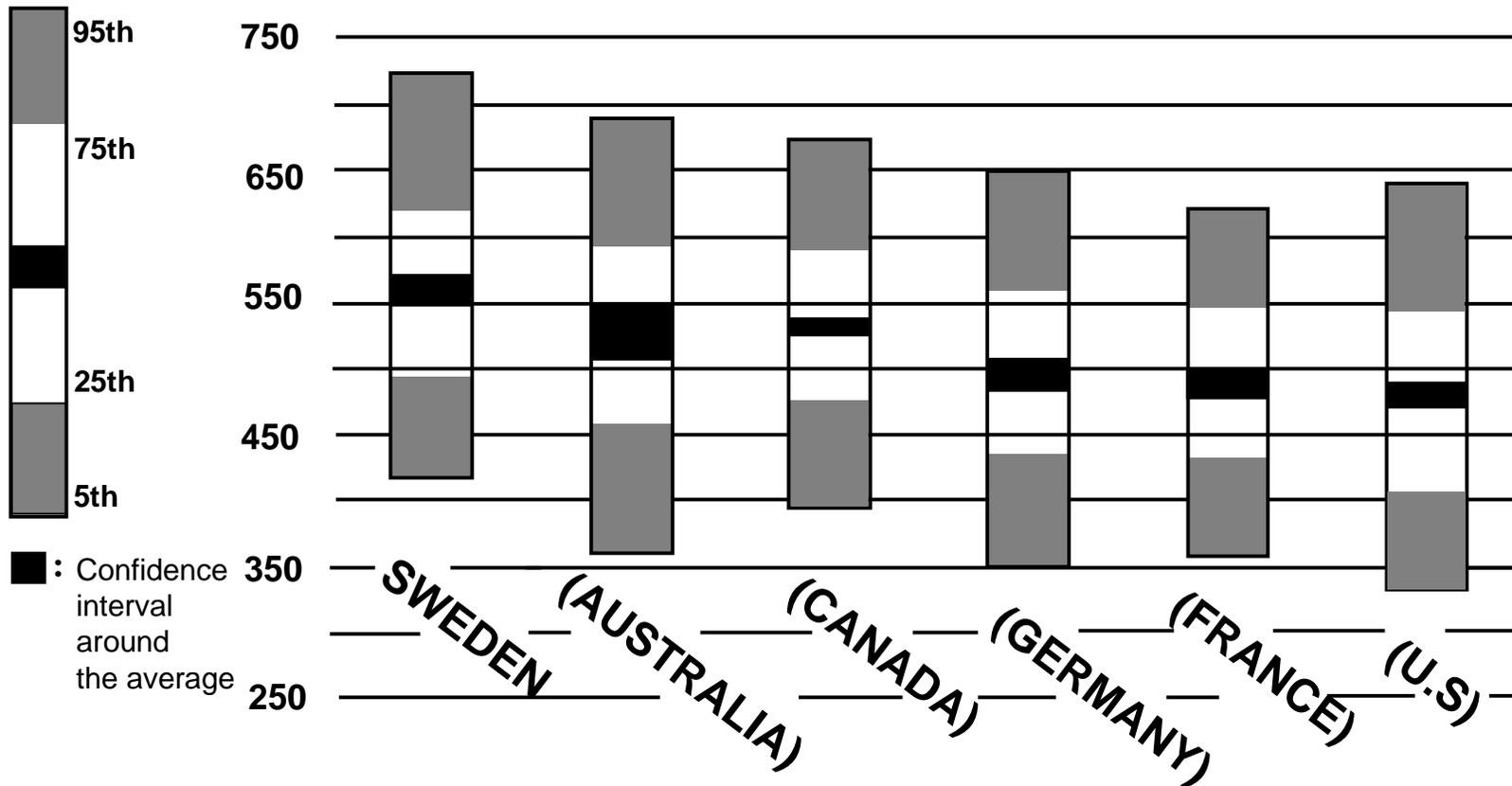
NATIONS WITH AVERAGE SCORES SIGNIFICANTLY LOWER THAN THE U.S.	
NATION	AVERAGE
(CYPRUS)	448
(SOUTH AFRICA)	349

NOTE: Nations not meeting international sampling and other guidelines are shown in parentheses.

SOURCE: Mullis et al. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School*.
Table 2.2. Chestnut Hill, MA: Boston College.

FINAL YEAR OF SECONDARY SCHOOL SCIENCE GENERAL KNOWLEDGE DISTRIBUTION OF SCORES FOR SELECTED COUNTRIES

PERCENTILES OF PERFORMANCE



NOTE: Nations not meeting international sampling and other guidelines are shown in parentheses.

SOURCE: Mullis et al. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School*. Tables 8.1 and E.5. Chestnut Hill, MA: Boston College.

FINAL YEAR OF
SECONDARY SCHOOL:
RELEASED ITEMS

On Boston College TIMSS Web site at:

<http://www.csteep.bc.edu/TIMSS1/Items.html>

EXAMPLE 1: SCIENCE GENERAL KNOWLEDGE ITEM

Correct answer examples:

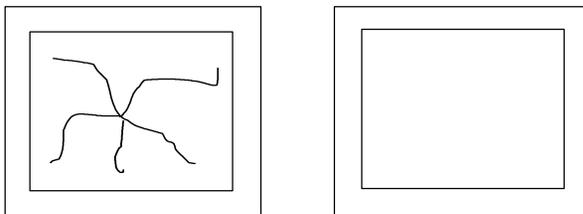
- *“Drinking from the same cup”*
- *“Shaking hands”*
- *Stating that he got it from someone who had the flu.*

*U.S. Average: **59 percent***

*International Average: **68 percent***

EXAMPLE 2: SCIENCE GENERAL KNOWLEDGE ITEM

The sketch below shows two windows. The left window has been cracked by a flying stone. A tennis ball, with the same mass and speed as the stone, strikes the adjacent, similar window, but does not crack it.



What is one important reason why the impact of the stone cracks the window but the impact of the tennis ball does not?

Correct answer examples:

- *“The rock is hard and penetrates the glass causing the crack, but the ball is soft and when it strikes the window it absorbs some of its own velocity and rebounds back.”*
- *“The surface area struck is greater on the tennis ball than on the stone.”*

*U.S. Average: **54 percent***

*International Average: **57 percent***

EXAMPLE 3: SCIENCE GENERAL KNOWLEDGE ITEM

Electrical energy is used to power a lamp.

Is the amount of light energy produced more than, less than, or the same as the amount of electrical energy used?

The amount of light energy is: (check one)

----- more than

----- less than

----- the same as

the amount of electrical energy used. Give a reason to support your answer.

Correct answer examples:

- *“Less, because (much) energy is transformed to heat”.*
- *“Less, because energy (heat) is lost to the surroundings”.*

U.S. Average: 11 percent

International Average: 21 percent

FINAL YEAR OF SECONDARY SCHOOL

AVERAGE ADVANCED MATHEMATICS PERFORMANCE OF ADVANCED MATHEMATICS STUDENTS IN ALL COUNTRIES

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY HIGHER THAN THE U.S.	
NATION	AVERAGE
FRANCE	557
(RUSSIAN FEDERATION)	542
SWITZERLAND	533
(AUSTRALIA)	525
(DENMARK)	522
(CYPRUS)	518
(LITHUANIA)	516
GREECE	513
SWEDEN	512
CANADA	509
(SLOVENIA)	475

NATIONS WITH AVERAGE SCORES NOT SIGNIFICANTLY DIFFERENT FROM THE U.S.	
NATION	AVERAGE
(ITALY)	474
CZECH REPUBLIC	469
(GERMANY)	465
(UNITED STATES)	442
(AUSTRIA)	436

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY LOWER THAN THE U.S.	
NATION	AVERAGE
NONE	

INTERNATIONAL AVERAGE=501

NOTE: Nations not meeting international sampling or other guidelines are shown in parentheses.

SOURCE: Mullis et al. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School*.
Table 5.1. Chestnut Hill, MA: Boston College.

FINAL YEAR OF SECONDARY SCHOOL

AVERAGE ADVANCED MATHEMATICS PERFORMANCE OF ADVANCED MATHEMATICS STUDENTS IN OTHER COUNTRIES COMPARED WITH U.S. AP CALCULUS STUDENTS

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY HIGHER THAN THE U.S.	
NATION	AVERAGE
FRANCE	557

NATIONS WITH AVERAGE SCORES NOT SIGNIFICANTLY DIFFERENT FROM THE U.S.	
NATION	AVERAGE
(RUSSIAN FEDERATION)	542
SWITZERLAND	533
(AUSTRALIA)	525
(DENMARK)	522
(CYPRUS)	518
(LITHUANIA)	516
(UNITED STATES)	513
GREECE	513
SWEDEN	512
CANADA	509

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY LOWER THAN THE U.S.	
NATION	AVERAGE
(SLOVENIA)	475
(ITALY)	474
CZECH REPUBLIC	469
(GERMANY)	465
(AUSTRIA)	436

INTERNATIONAL AVERAGE=505

NOTE: Nations not meeting international sampling and other guidelines are shown in parentheses.

SOURCE: Mullis et al. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School*. Table 5.1. Chestnut Hill, MA: Boston College and National Center for Education Statistics (1998). *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context*, Figure 10. Washington, DC.

NUMBER OF COUNTRIES WITH AVERAGE SCORES HIGHER THAN, SIMILAR TO, OR LOWER THAN THE UNITED STATES

FINAL YEAR OF SECONDARY SCHOOL ADVANCED MATHEMATICS CONTENT AREAS

COUNTRIES	NUMBERS & EQUATIONS	CALCULUS	GEOMETRY
HIGHER THAN THE U.S.	11	11	15
SIMILAR TO THE U.S.	3	3	0
LOWER THAN THE U.S.	1	1	0

SOURCE: National Center for Education Statistics (1998). *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context*, Figure 12. Washington, D.C.

FINAL YEAR OF SECONDARY SCHOOL

AVERAGE PHYSICS PERFORMANCE OF ADVANCED SCIENCE STUDENTS IN ALL COUNTRIES

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY HIGHER THAN THE U.S.	
NATION	AVERAGE
NORWAY	581
SWEDEN	573
(RUSSIAN FEDERATION)	545
(DENMARK)	534
(SLOVENIA)	523
(GERMANY)	522
(AUSTRALIA)	518
(CYPRUS)	494
(LATVIA)	488
SWITZERLAND	488
GREECE	486
(CANADA)	485
FRANCE	466
CZECH REPUBLIC	451

NATIONS WITH AVERAGE SCORES NOT SIGNIFICANTLY DIFFERENT FROM THE U.S.	
NATION	AVERAGE
(AUSTRIA)	435
(UNITED STATES)	423

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY LOWER THAN THE U.S.	
NATION	AVERAGE
NONE	

INTERNATIONAL AVERAGE=501

NOTE: Nations not meeting international sampling and other guidelines are shown in parentheses.

SOURCE: Mullis et al. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School*. Table 8.1. Chestnut Hill, MA: Boston College.

FINAL YEAR OF SECONDARY SCHOOL

AVERAGE PHYSICS PERFORMANCE OF ADVANCED SCIENCE STUDENTS IN OTHER COUNTRIES COMPARED WITH U.S. AP PHYSICS STUDENTS

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY HIGHER THAN THE U.S.	
NATION	AVERAGE
NORWAY	581
SWEDEN	573
(RUSSIAN FEDERATION)	545
(DENMARK)	534
(GERMANY)	522
(AUSTRALIA)	518

INTERNATIONAL AVERAGE=504

NATIONS WITH AVERAGE SCORES NOT SIGNIFICANTLY DIFFERENT FROM THE U.S.	
NATION	AVERAGE
(SLOVENIA)*	523
(CYPRUS)	494
(LATVIA)	488
SWITZERLAND	488
GREECE	486
(CANADA)	485
(UNITED STATES)	474
FRANCE	466
CZECH REPUBLIC	451

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY LOWER THAN THE U.S.	
NATION	AVERAGE
(AUSTRIA)	435

NOTES:

Nations not meeting international sampling and other guidelines are shown in parentheses.

* The placement of Slovenia may appear to be out of place; however, statistically, its placement is correct.

SOURCES: Mullis et al. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School*. Table 8.1. Chestnut Hill, MA: Boston College and National Center for Education Statistics (1998). *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context*, Figure 17. Washington, DC.

NUMBER OF COUNTRIES WITH AVERAGE SCORES HIGHER THAN, SIMILAR TO, OR LOWER THAN THE UNITED STATES

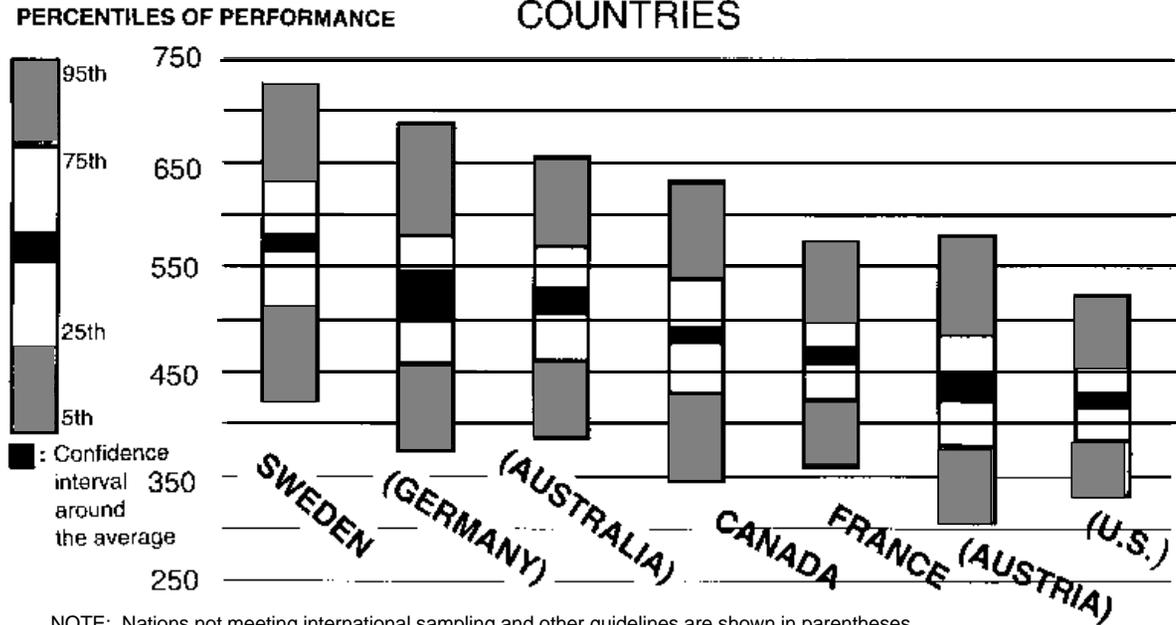
FINAL YEAR OF SECONDARY SCHOOL PHYSICS CONTENT AREAS

COUNTRIES	MECHANICS	ELECTRICITY & MAGNETISM	HEAT	WAVE PHENOMENA	MODERN PHYSICS
HIGHER THAN THE U.S.	14	14	9	10	11
SIMILAR TO THE U.S.	1	1	5	5	3
LOWER THAN THE U.S.	0	1	1	0	1

SOURCE: National Center for Education Statistics (1998). *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context*, Figure 18. Washington, D.C.

FINAL YEAR OF SECONDARY SCHOOL PHYSICS

DISTRIBUTION OF SCORES FOR SELECTED COUNTRIES



NOTE: Nations not meeting international sampling and other guidelines are shown in parentheses.

SOURCE: Mullis et al. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School*. Tables 8.1 and E.5. Chestnut Hill, MA: Boston College.

PHYSICS EXAMPLE 1: HEAT ITEM

When a small volume of water is boiled, a large volume of steam is produced. Why?

- A. The molecules are further apart in steam than in water.
- B. Water molecules expand when heated.
- C. The change from water to steam causes the number of molecules to increase.
- D. Atmospheric pressure works more on water molecules than on steam molecules.
- E. Water molecules repel each other when heated.

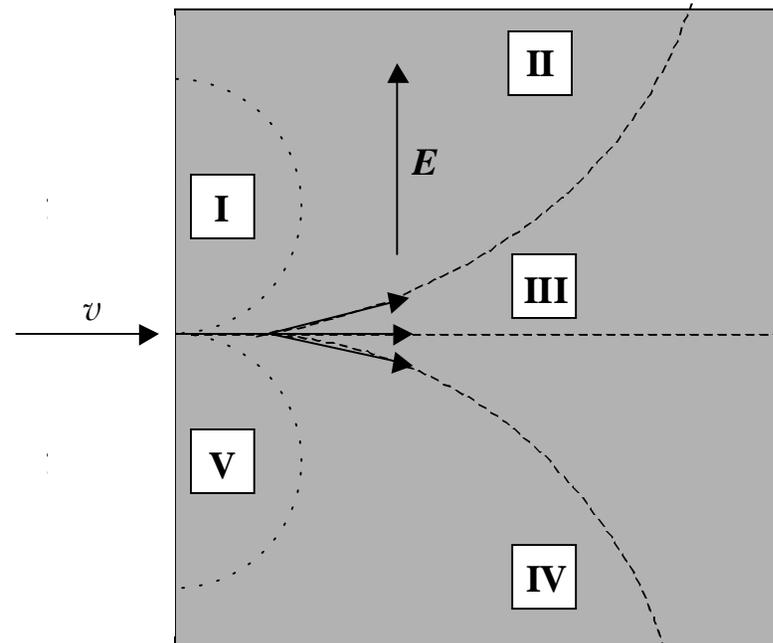
*Correct Answer: A U.S. Average: **60 percent** International Average: **65 percent***

PHYSICS EXAMPLE 2: ELECTRICITY AND MAGNETISM ITEM

Electrons enter a uniform electric field E with a velocity v as shown in the figure. The velocity v is perpendicular to the electric field E .

Which one of the dashed paths (I, II, III, IV, or V) best represents the path of the electrons in the electric field?

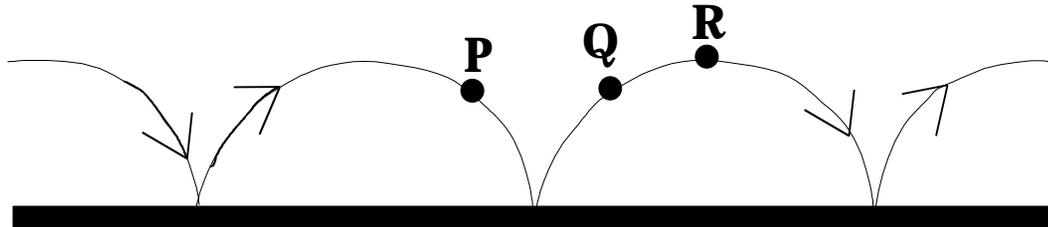
- A. I
- B. II
- C. III
- D. IV
- E. V



Correct Answer: D U.S. Average: 12 percent International Average: 32 percent

PHYSICS EXAMPLE 3: MECHANICS ITEM

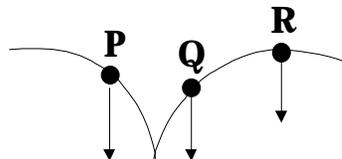
The figure shows the trajectory of a ball bouncing on a floor, with negligible air resistance.



Draw arrows on the figure showing the direction of the acceleration of the ball at points P, Q, and R.

Correct Answer:

The acceleration is parallel to g , downwards arrows at P, Q and R.



U.S. Average: 6 percent

International Average: 16 percent

FOURTH GRADE

SCIENCE ACHIEVEMENT

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY HIGHER THAN THE U.S.	
NATION	AVERAGE
KOREA	597

NATIONS WITH AVERAGE SCORES NOT SIGNIFICANTLY DIFFERENT FROM THE U.S.	
NATION	AVERAGE
JAPAN	574
UNITED STATES	565
(AUSTRIA)	565
(AUSTRALIA)	562
(NETHERLANDS)	557
CZECH REPUBLIC	557

INTERNATIONAL AVERAGE=524

SOURCE: Martin et al. (1997) *Science Achievement in the Primary School Years*. Table 1.1. Boston College: Chestnut Hill, MA.

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY LOWER THAN THE U.S.	
NATION	AVERAGE
(ENGLAND)	551
CANADA	549
SINGAPORE	547
(SLOVENIA)	546
IRELAND	539
SCOTLAND	536
HONG KONG	533
(HUNGARY)	532
NEW ZEALAND	531
NORWAY	530
(LATVIA (LSS))	512
(ISRAEL)	505
ICELAND	505
GREECE	497
PORTUGAL	480
CYPRUS	475
(THAILAND)	473
IRAN, ISLAMIC REPUBLIC	416
(KUWAIT)	401

NOTES:

1. Nations not meeting international sampling and other guidelines are shown in parentheses.
2. Latvia is designated LSS because only Latvian-speaking schools were tested, which represents less than 65 percent of the population.
3. The international average is the average of the national averages of the 26 nations.

EIGHTH GRADE

SCIENCE ACHIEVEMENT

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY HIGHER THAN THE U.S.	
NATION	AVERAGE
SINGAPORE	607
CZECH REPUBLIC	574
JAPAN	571
KOREA	565
(BULGARIA)	565
(NETHERLANDS)	560
(SLOVENIA)	560
(AUSTRIA)	558
HUNGARY	554

NATIONS WITH AVERAGE SCORES NOT SIGNIFICANTLY DIFFERENT FROM THE U.S.	
NATION	AVERAGE
(ENGLAND)	552
BELGIUM-FLEMISH	550
(AUSTRALIA)	545
SLOVAK REPUBLIC	544
RUSSIAN FEDERATION	538
IRELAND	538
SWEDEN	535
UNITED STATES	534
(GERMANY)	531
CANADA	531
NORWAY	527
NEW ZEALAND	525
(THAILAND)	525
(ISRAEL)	524
HONG KONG	522
SWITZERLAND	522
(SCOTLAND)	517

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY LOWER THAN THE U.S.	
NATION	AVERAGE
SPAIN	517
FRANCE	498
(GREECE)	497
ICELAND	494
(ROMANIA)	486
(LATVIA (LSS))	485
PORTUGAL	480
(DENMARK)	478
(LITHUANIA)	476
(BELGIUM-FRENCH)	471
IRAN, ISLAMIC REPUBLIC	470
CYPRUS	463
(KUWAIT)	430
(COLOMBIA)	411
(SOUTH AFRICA)	326

INTERNATIONAL AVERAGE=516

SOURCE: Beaton et al. (1996). *Science Achievement in the Middle School Years*. Table 1.1. Boston College: Chestnut Hill, MA.

NOTES:

1. Nations not meeting international sampling and other guidelines are shown in parentheses.
2. Latvia is designated LSS because only Latvian-speaking schools were tested, which represents less than 65 percent of the population.
3. The international average is the average of the national averages of the 41 nations.
4. The country average of Scotland (or Spain) may appear to be out of place; however, statistically, its placement is correct.

FINAL YEAR OF SECONDARY SCHOOL

SCIENCE GENERAL KNOWLEDGE ACHIEVEMENT

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY HIGHER THAN THE U.S.	
NATION	AVERAGE
SWEDEN	559
(NETHERLANDS)	558
(ICELAND)	549
(NORWAY)	544
(CANADA)	532
NEW ZEALAND	529
(AUSTRALIA)	527
SWITZERLAND	523
(AUSTRIA)	520
(SLOVENIA)	517
(DENMARK)	509

NATIONS WITH AVERAGE SCORES NOT SIGNIFICANTLY DIFFERENT FROM THE U.S.	
NATION	AVERAGE
(GERMANY)	497
(FRANCE)	487
CZECH REPUBLIC	487
(RUSSIAN FEDERATION)	481
(UNITED STATES)	480
(ITALY)	475
HUNGARY	471
(LITHUANIA)	461

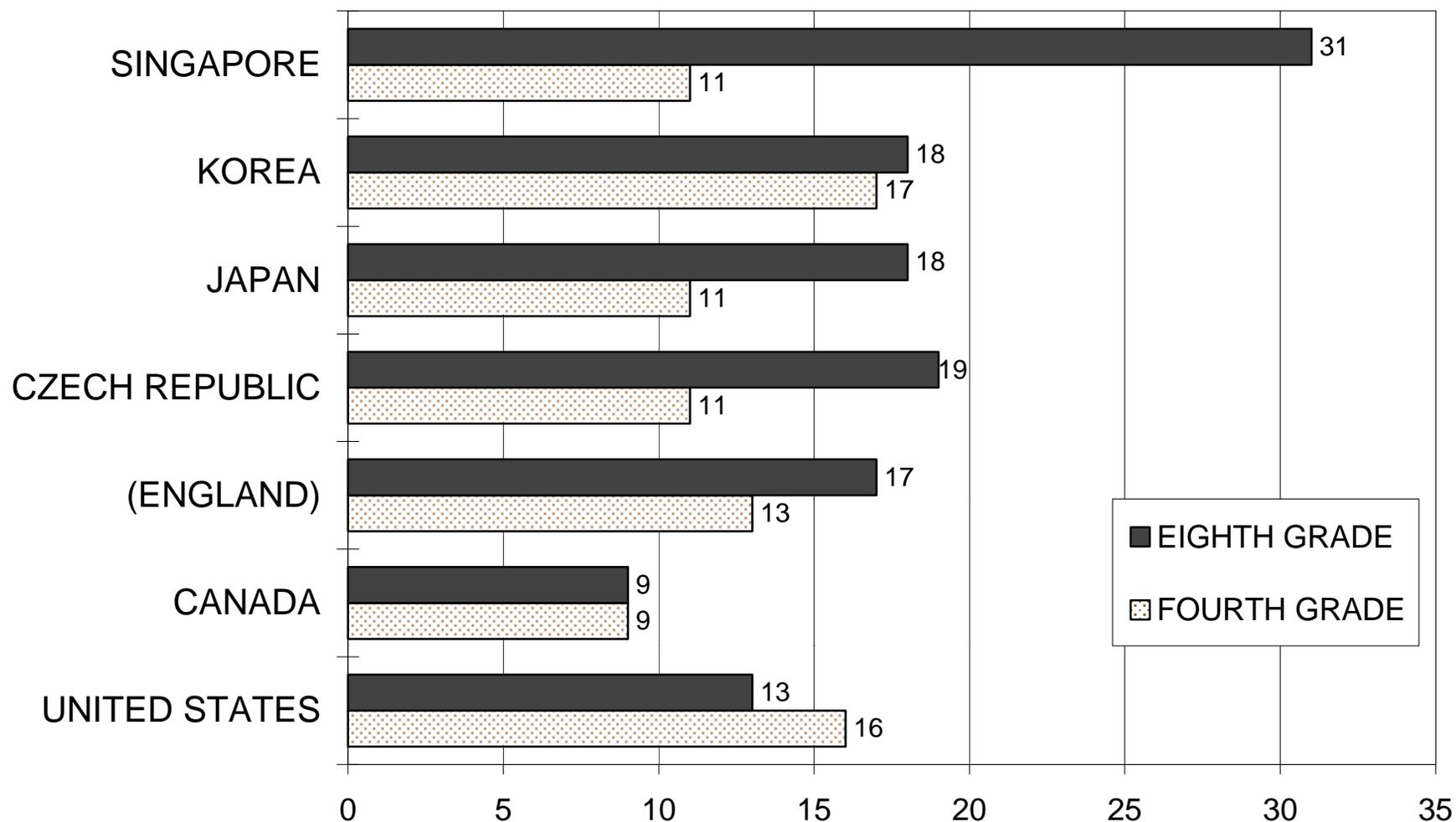
NATIONS WITH AVERAGE SCORES SIGNIFICANTLY LOWER THAN THE U.S.	
NATION	AVERAGE
(CYPRUS)	448
(SOUTH AFRICA)	349

INTERNATIONAL AVERAGE=500

NOTE: Nations not meeting international sampling and other guidelines are shown in parentheses.

SOURCE: Mullis et al. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School*. Table 2.2. Chestnut Hill, MA: Boston College.

PERCENTAGE OF STUDENTS FROM SELECTED NATIONS SCORING AMONG THE TOP TEN PERCENT OF FOURTH AND EIGHTH GRADERS IN SCIENCE

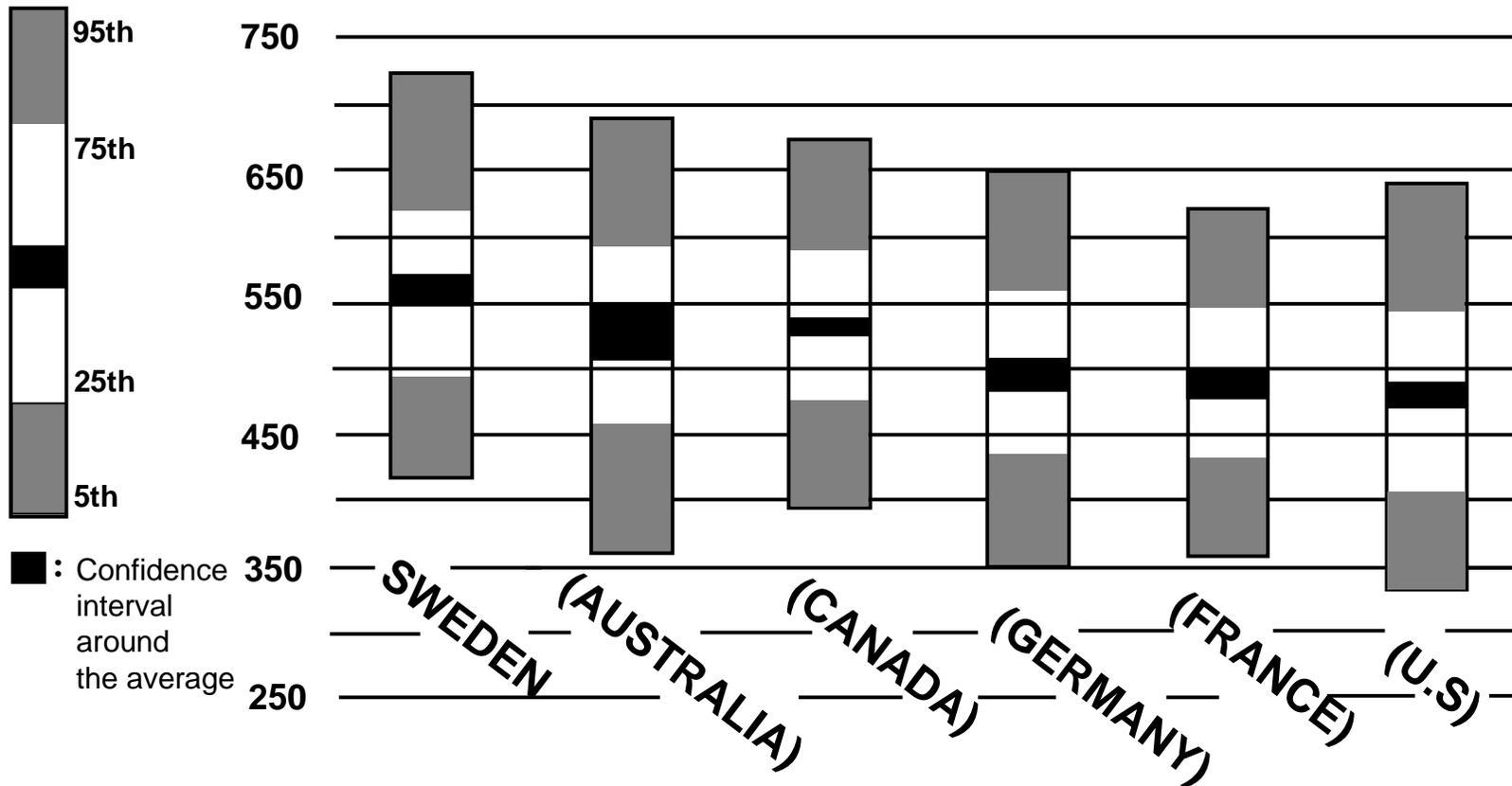


SOURCE: National Center for Education Statistics (1997). *Pursuing Excellence: A Study of U.S. Fourth-Grade Mathematics and Science Achievement in International Context*, Figure 4, and *Pursuing Excellence: A Study of U.S. Eighth-Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context*. Figure 4. Washington, D.C.

NOTE: England did not meet international sampling guidelines at the fourth and eighth grades.

FINAL YEAR OF SECONDARY SCHOOL SCIENCE GENERAL KNOWLEDGE DISTRIBUTION OF SCORES FOR SELECTED COUNTRIES

PERCENTILES OF PERFORMANCE



NOTE: Nations not meeting international sampling and other guidelines are shown in parentheses.

SOURCE: Mullis et al. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School*. Tables 8.1 and E.5. Chestnut Hill, MA: Boston College.

FINAL YEAR OF SECONDARY SCHOOL

AVERAGE PHYSICS PERFORMANCE OF ADVANCED SCIENCE STUDENTS IN OTHER COUNTRIES COMPARED WITH U.S. AP PHYSICS STUDENTS

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY HIGHER THAN THE U.S.	
NATION	AVERAGE
NORWAY	581
SWEDEN	573
(RUSSIAN FEDERATION)	545
(DENMARK)	534
(GERMANY)	522
(AUSTRALIA)	518

INTERNATIONAL AVERAGE=504

NATIONS WITH AVERAGE SCORES NOT SIGNIFICANTLY DIFFERENT FROM THE U.S.	
NATION	AVERAGE
(SLOVENIA)*	523
(CYPRUS)	494
(LATVIA)	488
SWITZERLAND	488
GREECE	486
(CANADA)	485
(UNITED STATES)	474
FRANCE	466
CZECH REPUBLIC	451

NATIONS WITH AVERAGE SCORES SIGNIFICANTLY LOWER THAN THE U.S.	
NATION	AVERAGE
(AUSTRIA)	435

NOTES:

Nations not meeting international sampling and other guidelines are shown in parentheses.

* The placement of Slovenia may appear to be out of place; however, statistically, its placement is correct.

SOURCES: Mullis et al. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School*. Table 8.1. Chestnut Hill, MA: Boston College and National Center for Education Statistics (1998). *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context*, Figure 17. Washington, DC.

NUMBER OF COUNTRIES WITH AVERAGE SCORES HIGHER THAN, SIMILAR TO, OR LOWER THAN THE UNITED STATES

FOURTH-GRADE SCIENCE CONTENT AREAS

COUNTRIES	EARTH SCIENCE	LIFE SCIENCE	PHYSICAL SCIENCE	ENVIRONMENTAL ISSUES
HIGHER THAN THE U.S.	2	1	5	1
SIMILAR TO THE U.S.	8	6	7	2
LOWER THAN THE U.S.	15	18	13	22

SOURCE: National Center for Education Statistics (1997) *Pursuing Excellence: A Study of U.S. Fourth-Grade Mathematics and Science Achievement in International Context*, Figure 6. Washington, D.C.

NUMBER OF COUNTRIES WITH AVERAGE SCORES HIGHER THAN, SIMILAR TO, OR LOWER THAN THE UNITED STATES

EIGHTH-GRADE SCIENCE CONTENT AREAS

COUNTRIES	EARTH SCIENCE	LIFE SCIENCE	PHYSICS	CHEMISTRY	ENVIRONMENTAL ISSUES
HIGHER THAN THE U.S.	4	4	13	7	1
SIMILAR TO THE U.S.	22	21	17	21	14
LOWER THAN THE U.S.	14	15	10	12	25

SOURCE: National Center for Education Statistics (1997). *Pursuing Excellence: A Study of U.S. Eighth-Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context*, Figure 6. Washington, D.C.