Science Performance in the United States and Internationally

Average scores in 2011

In science, the U.S. national average score was 544 at grade 4 and 525 at grade 8 (tables 26 and 27). Both scores were higher than the TIMSS scale average of 500 at both grades.

Among the 45 *countries* that participated at grade 4, the U.S. average science score was among the top 6 (5 countries had higher average scores than the United States). Thirty-nine countries had lower average scores than the United States.

Among all 57 *education systems* that participated at grade 4 (i.e., both countries and other education systems, including U.S. states that participated in TIMSS with individual state samples), the United States was among the top 10 in average science scores (6 education systems had higher averages and 3 were not measurably different). Korea, Singapore, Finland, Japan, the Russian Federation, and Chinese Taipei-CHN had higher average scores than the United States; and Florida-USA, Alberta-CAN, and North Carolina-USA had average scores not measurably different from the U.S. average at grade 4. The United States outperformed 47 education systems.

At grade 8, among the 38 *countries* that participated in TIMSS, the U.S. average science score was among the top 10 (6 countries had higher averages and 3 had averages not measurably different from the United States). Twenty-eight countries had lower average scores than the United States.

Among all 56 *education systems* that participated at grade 8, the United States was among the top 23 education systems in average science scores (12 education systems had higher averages and 10 were not measurably different). Singapore, Massachusetts-USA, Chinese Taipei-CHN, Korea, Japan, Minnesota-USA, Finland, Alberta-CAN, Slovenia, the Russian Federation, Colorado-USA, and Hong Kong-CHN had higher average scores than the United States; and England-GBR, Indiana-USA, Connecticut-USA, North Carolina-USA, Florida-USA, Hungary, Ontario-CAN, Quebec-CAN, Australia, and Israel had average scores not measurably different from the U.S. average at grade 8. The United States had higher average science scores than 33 education systems.

Change in scores

Several education systems that participated in TIMSS 2011 also participated in the last administration of TIMSS in 2007 or in the first administration of TIMSS in 1995. Some education systems participated in both of these previous administrations. Comparing scores between previous administrations of TIMSS and the most recent administration provides perspective on change over time.[17]

Change at grade 4 between 2007 and 2011

Among the 28 education systems that participated in both the 2007 and 2011 TIMSS science assessments at grade 4, the average science score increased in 9 education systems and decreased in 5 education systems (figure 9). In the rest, including the United States, there was no measurable change in the average grade 4 science scores between 2007 and 2011.

The education systems in which 4th-graders' average scores increased between 2007 and 2011 were Georgia (37 points), Tunisia (27 points), the Czech Republic (21 points), Norway (17 points), the Islamic Republic of Iran (17 points), Denmark (11 points), Japan (11 points), Sweden (9 points), and the Netherlands (8 points). None of these increases changed these education systems' standing relative to the United States between 2007 and 2011.[18]

Scores decreased at grade 4 during this time in Hong Kong-CHN (19 points), England-GBR (13 points), Australia (12 points), Italy (11 points), and New Zealand (7 points). As a result, U.S. average performance at grade 4 went from below the average of Hong Kong-CHN in 2007 to higher than that country's average in 2011, and from not measurably different from the averages of England-GBR and Italy in 2007 to higher than their averages in 2011.[19]

[17]Several participating countries that are reported with the 2011 results in other tables in this report are excluded from these comparisons based on International Study Center (ISC) review of assessment results. Morocco and Yemen participated in both 2007 or 1995 and 2011 at grade 4, but had unreliable 2011 science scores. Armenia, Kazakhstan, and Qatar also participated in both 2007 and 2011 at grade 4, but their 2007 science scores were not comparable to their 2011 scores. Armenia, Qatar, Saudi Arabia, and Turkey participated in both 2007 and 2011 at grade 8, but their 2007 science scores were not comparable to their 2011 scores. Similarly, Italy, Kuwait, and Thailand participated in both 1995 and 2011 at grade 4 and 8, but their 1995 science scores were not comparable to their 2011 scores.

[18]Although the average score of the Russian Federation did not increase measurably, its standing relative to the United States moved from being not measurably different in 2007 to scoring above the United States in 2011.

[19]Although the average score of Hungary and Ontario-CAN did not decrease measurably, their standing relative to the United States moved from being not measurably different in 2007 to scoring below the United States in 2011.

Table 26. Average science scores of 4th-grade students, by education system: 2011

Grade 4	
Education system	**Average score**
TIMSS scale average	500
Korea, Rep. of	587 ▲
Singapore[1]	583 ▲
Finland	570 ▲
Japan	559 ▲
Russian Federation	552 ▲
Chinese Taipei-CHN	552 ▲
United States[1]	**544**
Czech Republic	536 ▼
Hong Kong-CHN[1]	535 ▼
Hungary	534 ▼
Sweden	533 ▼
Slovak Republic	532 ▼
Austria	532 ▼
Netherlands[2]	531 ▼
England-GBR	529 ▼
Denmark[1]	528 ▼
Germany	528 ▼
Italy	524 ▼
Portugal	522 ▼
Slovenia	520 ▼
Northern Ireland-GBR[2]	517 ▼
Ireland	516 ▼
Croatia[1]	516 ▼
Australia	516 ▼
Serbia[1]	516 ▼
Lithuania[1,3]	515 ▼
Belgium (Flemish)-BEL	509 ▼
Romania	505 ▼
Spain	505 ▼
Poland	505 ▼

Grade 4	
Education system	**Average score**
New Zealand	497 ▼
Kazakhstan[1]	495 ▼
Norway[4]	494 ▼
Chile	480 ▼
Thailand	472 ▼
Turkey	463 ▼
Georgia[3,5]	455 ▼
Iran, Islamic Rep. of	453 ▼
Bahrain	449 ▼
Malta	446 ▼
Azerbaijan[1,5]	438 ▼
Saudi Arabia	429 ▼
United Arab Emirates	428 ▼
Armenia	416 ▼
Qatar[1]	394 ▼
Oman	377 ▼
Kuwait[3,6]	347 ▼
Tunisia[6]	346 ▼
Morocco[7]	264 ▼
Yemen[7]	209 ▼
Benchmarking education systems	
Florida-USA[3,8]	545
Alberta-CAN[1]	541
North Carolina-USA[1,3]	538
Ontario-CAN	528 ▼
Quebec-CAN	516 ▼
Dubai-UAE	461 ▼
Abu Dhabi-UAE	411 ▼

▲ Average score is higher than U.S. average score.
▼ Average score is lower than U.S. average score.
[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Met guidelines for sample participation rates only after replacement schools were included.
[3]National Target Population does not include all of the International Target Population (see appendix A).
[4]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[5]Exclusion rates for Azerbaijan and Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[7]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
[8]National Defined Population covers less than 90 percent, but at least 77 percent of National Target Population (see appendix A).
NOTE: Education systems are ordered by 2011 average score. Italics indicate participants identified and counted in this report as an education system and not as a separate country. Participants that did not administer TIMSS at the target grade are not shown; see the international report for their results. All U.S. state data are based on public school students only. All average scores reported as higher or lower than the U.S. average score are different at the .05 level of statistical significance. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one education system may not be significant while a large difference between the United States and another education system may not be significant. The standard errors of the estimates are shown in table E-22 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Table 27. Average science scores of 8th-grade students, by education system: 2011

Education system	Grade 8 Average score
TIMSS scale average	500
Singapore[1]	590 ▲
Chinese Taipei-CHN	564 ▲
Korea, Rep. of	560 ▲
Japan	558 ▲
Finland	552 ▲
Slovenia	543 ▲
Russian Federation[1]	542 ▲
Hong Kong-CHN	535 ▲
England-GBR[2]	533
United States[1]	**525**
Hungary	522
Australia	519
Israel[3]	516
Lithuania[4]	514 ▼
New Zealand	512 ▼
Sweden	509 ▼
Italy	501 ▼
Ukraine	501 ▼
Norway	494 ▼
Kazakhstan	490 ▼
Turkey	483 ▼
Iran, Islamic Rep. of	474 ▼
Romania	465 ▼
United Arab Emirates	465 ▼
Chile	461 ▼
Bahrain	452 ▼
Thailand	451 ▼
Jordan	449 ▼
Tunisia	439 ▼
Armenia	437 ▼
Saudi Arabia	436 ▼
Malaysia	426 ▼
Syrian Arab Republic	426 ▼
Palestinian Nat'l Auth.	420 ▼
Georgia[4,5]	420 ▼
Oman	420 ▼
Qatar	419 ▼
Macedonia, Rep. of	407 ▼
Lebanon	406 ▼
Indonesia	406 ▼
Morocco	376 ▼
Ghana[6]	306 ▼
Benchmarking education systems	
Massachusetts-USA[1,4]	567 ▲
Minnesota-USA[4]	553 ▲
Alberta-CAN[1]	546 ▲
Colorado-USA[4]	542 ▲
Indiana-USA[1,4]	533
Connecticut-USA[1,4]	532
North Carolina-USA[3,4]	532
Florida-USA[1,4]	530
Ontario-CAN[1]	521
Quebec-CAN	520
California-USA[1,4]	499 ▼
Alabama-USA[4]	485 ▼
Dubai-UAE	485 ▼
Abu Dhabi-UAE	461 ▼

▲ Average score is higher than U.S. average score.
▼ Average score is lower than U.S. average score.
[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[3]National Defined Population covers less than 90 percent, but at least 77 percent of National Target Population (see appendix A).
[4]National Target Population does not include all of the International Target Population (see appendix A).
[5]Exclusion rates for Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
NOTE: Education systems are ordered by 2011 average score. Italics indicate participants identified and counted in this report as an education system and not as a separate country. Participants that did not administer TIMSS at the target grade are not shown; see the international report for their results. All U.S. state data are based on public school students only. All average scores reported as higher or lower than the U.S. average score are different at the .05 level of statistical significance. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one education system may be significant while a large difference between the United States and another education system may not be significant. The standard errors of the estimates are shown in table E-23 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Change at grade 4 between 1995 and 2011

Among the 20 education systems that participated in both the 1995 and 2011 TIMSS science assessments at grade 4, the average science score increased in 9 education systems and decreased in 2 (figure 9). In the other 9 education systems that participated in TIMSS in both years, including the United States, there was no measurable change in the average grade 4 science scores between 1995 and 2011.

The education systems in which 4th-graders' average scores increased between 1995 and 2011 were the Islamic Republic of Iran (73 points), Portugal (70 points), Singapore (60 points), Slovenia (56 points), Hong Kong-CHN (27 points), Hungary (27 points), Ontario-CAN (11 points), Korea (11 points), and Japan (5 points). The increase in the Singapore average meant that it moved from having a lower average score at grade 4 than the United States in 1995 to having a higher average score in 2011.[20] The increases in the other education systems did not change their standing relative to the United States.

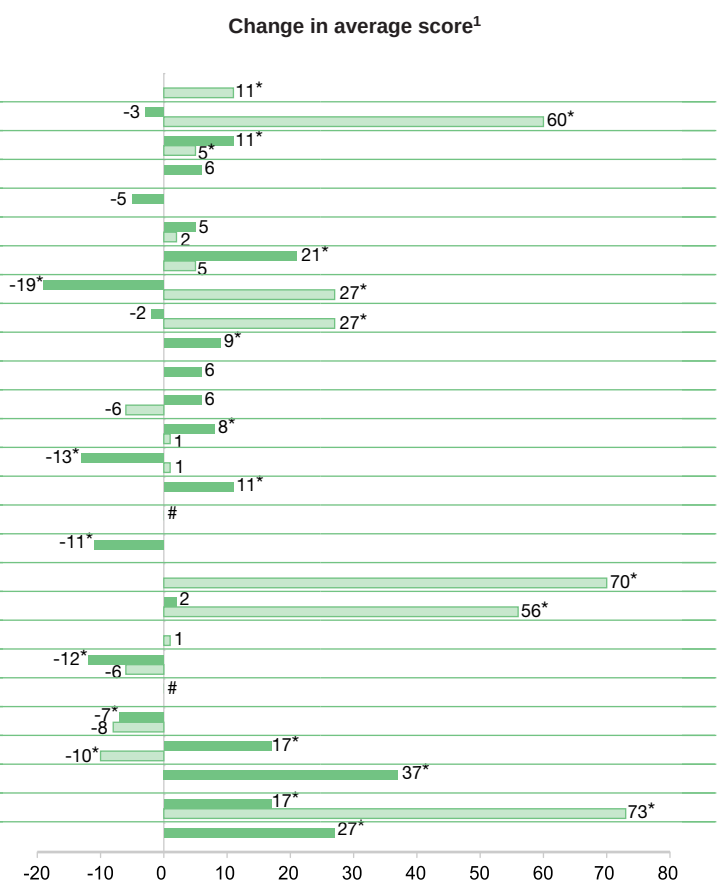
Scores decreased during this time for 4th-graders in Quebec-CAN (12 points) and Norway (10 points). These decreases did not change their standing relative to the United States.[21]

[20]Two-thirds of Singapore's increase (42 points) occurred between 1995 and 2003.

[21]Although the average score of Austria and New Zealand did not decrease measurably, their standing relative to the United States moved from being not measurably different in 1995 to scoring below the United States in 2011.

Figure 9. Change in average science scores of 4th-grade students, by education system: 2007–2011 and 1995–2011

Grade 4			
Average score			
1995	2007	2011	Education system
576 ▲		587 ▲	Korea, Rep. of
523 ▼	587 ▲	583 ▲	Singapore[2]
553 ▲	548 ▲	559 ▲	Japan
	546	552 ▲	Russian Federation
	557 ▲	552 ▲	*Chinese Taipei-CHN*
542	**539**	**544**	**United States[2]**
532 ▼	515 ▼	536 ▼	Czech Republic
508 ▼	554 ▲	535 ▼	*Hong Kong-CHN[2]*
508 ▼	536	534 ▼	Hungary
	525 ▼	533 ▼	Sweden
	526 ▼	532 ▼	Slovak Republic
538	526 ▼	532 ▼	Austria
530 ▼	523 ▼	531 ▼	Netherlands[3]
528 ▼	542	529 ▼	*England-GBR*
	517 ▼	528 ▼	Denmark[2]
	528 ▼	528 ▼	Germany
	535	524 ▼	Italy
452 ▼		522 ▼	Portugal
464 ▼	518 ▼	520 ▼	Slovenia
515 ▼		516 ▼	Ireland
521 ▼	527 ▼	516 ▼	Australia
	514 ▼	515 ▼	Lithuania[2,4]
505 ▼	504 ▼	497 ▼	New Zealand
504 ▼	477 ▼	494 ▼	Norway[5]
	418 ▼	455 ▼	Georgia[4,6]
380 ▼	436 ▼	453 ▼	Iran, Islamic Rep. of
	318 ▼	346 ▼	Tunisia[7]



See notes on next page.

Change at grade 8 between 2007 and 2011

At grade 8, among the 35 education systems that participated in both the 2007 and 2011 TIMSS science assessments, the average science score increased in 9 education systems and decreased in 7 education systems (figure 10). In the rest, including the United States, there was no measurable change in the average grade 8 science scores between 2007 and 2011.

The education systems in which 8th-graders' average scores increased between 2007 and 2011 were Singapore (23 points), the Palestinian National Authority (16 points), Ukraine (16 points), the Islamic Republic of Iran (15 points), Minnesota-USA (15 points), Quebec-CAN (13 points), the Russian Federation (13 points), Norway (8 points), and Korea (7 points). The increase in Quebec-CAN meant that its 8th-graders' average performance went from below that of U.S. 8th-graders in 2007 to being not measurably different from that of U.S. 8th-graders in 2011. None of the other education systems' increases changed their standing relative to the United States between 2007 and 2011.[22]

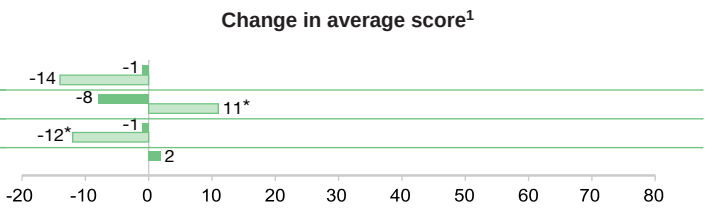
Scores decreased at grade 8 during this time in Malaysia (44 points), Jordan (33 points), the Syrian Arab Republic

[22]Although the average score of Hong Kong-CHN did not increase measurably, its standing relative to the United States moved from being not measurably different in 2007 to scoring above the United States in 2011.

Figure 9. Change in average science scores of 4th-grade students, by education system: 2007–2011 and 1995–2011—Continued

Grade 4

Average score			Benchmarking education systems
1995	2007	2011	
555	543	541	*Alberta-CAN*[2]
516 ▼	536	528 ▼	*Ontario-CAN*
529 ▼	517 ▼	516 ▼	*Quebec-CAN*
	460 ▼	461 ▼	*Dubai-UAE*



▲ Score is higher than U.S. score.
▼ Score is lower than U.S. score.
Change from 2007 to 2011.
Change from 1995 to 2011.
Rounds to zero.
*p<.05. Change in average scores is significant.
[1]The change in average score is calculated by subtracting the 2007 or 1995 estimate, respectively, from the 2011 estimate using unrounded numbers.
[2]National Defined Population covers 90 to 95 percent of National Target Population for 2011 (see appendix A).
[3]Met guidelines for sample participation rates only after replacement schools were included for 2011.
[4]National Target Population does not include all of the International Target Population for 2011 (see appendix A).
[5]Nearly satisfied guidelines for sample participation rates after replacement schools were included for 2011.
[6]Exclusion rates for Azerbaijan and Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available for 2011.
[7]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation in 2011 exceeds 15 percent, though it is less than 25 percent.
NOTE: Education systems are ordered by 2011 average scores. Italics indicate participants identified and counted in this report as an education system and not as a separate country. All education systems met international sampling and other guidelines in 2011, except as noted. Data are not shown for some education systems because comparable data from previous cycles are not available. Participants that did not administer TIMSS at the target grade are not shown; see the international report for their results. All U.S. state data are based on public school students only.
For 1995, Korea, Portugal, and Ontario-CAN had National Defined Population covering 90 to 95 percent of National Target Population; England-GBR had National Defined Population that covered less than 90 percent of National Target Population (but at least 77 percent); England-GBR, Netherlands, Australia, and Austria did not satisfy guidelines for sample participation rates. For 2007, the United States, Quebec-CAN, Ontario-CAN, and Alberta-CAN had National Defined Population covering 90 to 95 percent of National Target Population; the United States and Denmark met guidelines for sample participation rates only after replacement schools were included; the Netherlands and Dubai-UAE nearly satisfied guidelines for sample participation rates after replacement schools were included; Georgia had a National Target Population that did not include all of the International Target Population; Dubai-UAE tested the same cohort of students as other countries, but later in the assessment year at the beginning of the next school year.
All average scores reported as higher or lower than the U.S. average score are different at the .05 level of statistical significance. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between averages for one education system may be significant, while a large difference for another education system may not be significant. Detail may not sum to totals because of rounding. The standard errors of the estimates are shown in table E-24 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 2007, and 2011.

(26 points), Indonesia (21 points), Thailand (20 points), Hungary (17 points), and Bahrain (15 points). The decrease in Hungary's score meant that the U.S. average performance at grade 8 went from below the average in Hungary in 2007 to not measurably different from Hungary's score in 2011.[23]

[23]Although the average score of England-GBR and Lithuania did not decrease measurably, England-GBR's standing relative to the United States moved from above the United States in 2007 to being not measurably different in 2011, and Lithuania's standing relative to the United States moved from being not measurably different in 2007 to scoring below the United States in 2011.

Change at grade 8 between 1995 and 2011

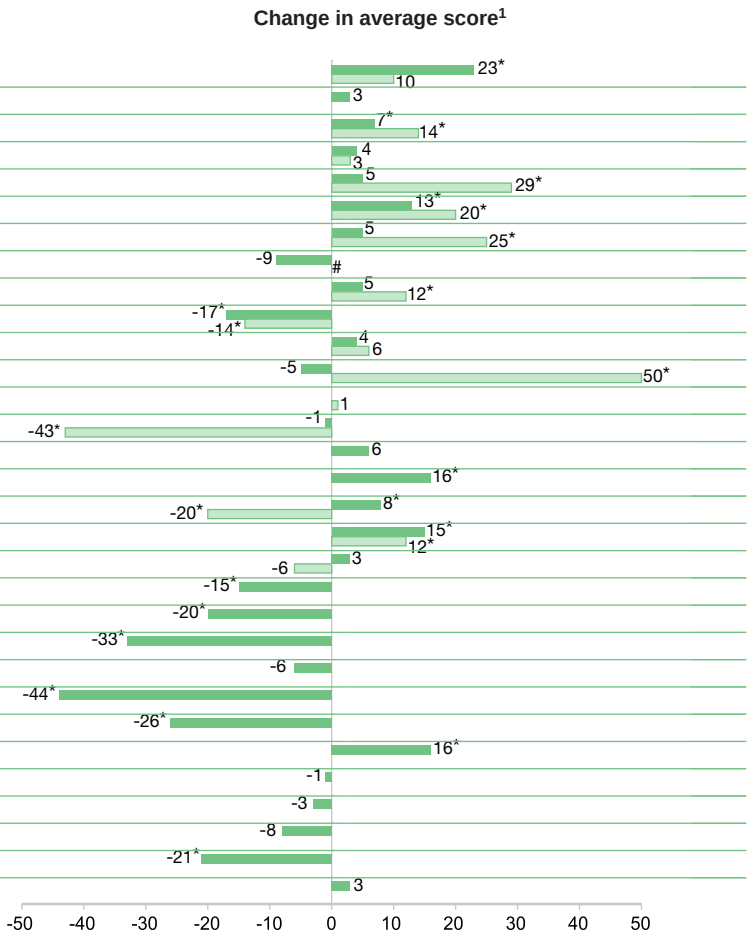
At grade 8, among the 20 education systems that participated in both the 1995 and 2011 TIMSS science assessments, the average science score increased in 8 education systems, including the United States, and decreased in 3 education systems (figure 10). In the rest, there was no measurable change in the average grade 8 science scores between 1995 and 2011.

The U.S. increase in average science score at grade 8 between 1995 and 2011 was 12 score points (from 513 to 525). Two education systems had larger increases than

Figure 10. Change in average science scores of 8th-grade students, by education system: 2007–2011 and 1995–2011

Grade 8

Average score 1995	Average score 2007	Average score 2011	Education system
580 ▲	567 ▲	590 ▲	Singapore[2]
	561 ▲	564 ▲	*Chinese Taipei-C HN*
546 ▲	553 ▲	560 ▲	Korea, Rep. of
554 ▲	554 ▲	558 ▲	Japan
514	538 ▲	543 ▲	Slovenia
523	530 ▲	542 ▲	Russian Federation[2]
510	530	535 ▲	*Hong Kong-CHN*
533 ▲	542 ▲	533	*England-GBR*[3]
513	**520**	**525**	**United States[2]**
537 ▲	539 ▲	522	Hungary
514	515	519	Australia
464 ▼	519	514 ▼	Lithuania[4]
511		512 ▼	New Zealand
553 ▲	511 ▼	509 ▼	Sweden
	495 ▼	501 ▼	Italy
	485 ▼	501 ▼	Ukraine
514	487 ▼	494 ▼	Norway
463 ▼	459 ▼	474 ▼	Iran, Islamic Rep. of
471 ▼	462 ▼	465 ▼	Romania
	467 ▼	452 ▼	Bahrain
	471 ▼	451 ▼	Thailand
	482 ▼	449 ▼	Jordan
	445 ▼	439 ▼	Tunisia
	471 ▼	426 ▼	Malaysia
	452 ▼	426 ▼	Syrian Arab Republic
	404 ▼	420 ▼	*Palestinian Nat'l Auth.*
	421 ▼	420 ▼	Georgia[4,5]
	423 ▼	420 ▼	Oman
	414 ▼	406 ▼	Lebanon
	427 ▼	406 ▼	Indonesia
	303 ▼	306 ▼	Ghana[6]



See notes on next page.

the United States during this time: Lithuania (50 points) and Slovenia (29 points). However, U.S. average performance at grade 8 went from being not measurably different than Slovenia, Hong Kong-CHN, and the Russian Federation in 1995 to having a lower average score in 2011; and from being above that of Ontario-CAN in 1995 to being not measurably different in 2011. The increases in the other education systems did not change their standing relative to the United States.

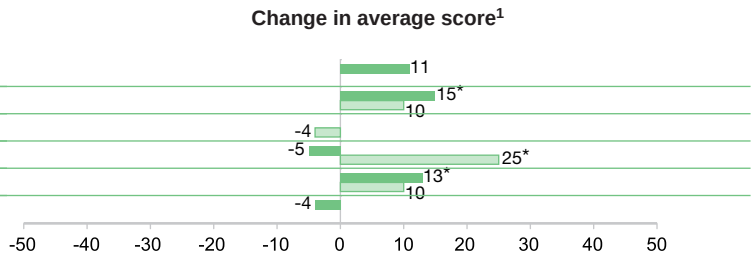
Scores decreased during this time at grade 8 in Sweden (43 points), Norway (20 points), and Hungary (14 points). As a result, Sweden moved from having a higher average score than the United States at grade 8 in 1995 to having a lower average score in 2011, Hungary moved from having a higher average score than the United States in 1995 to being not measurably different in 2011, and Norway moved from being not measurably different from the United States in 1995 to having a lower average score in 2011.[24]

[24]Although the average score of England-GBR and New Zealand did not decrease measurably, England-GBR's standing relative to the United States moved from above the United States in 1995 to being not measurably different in 2011, and New Zealand's standing relative to the United States moved from being not measurably different in 1995 to scoring below the United States in 2011.

Figure 10. Change in average science scores of 8th-grade students, by education system: 2007–2011 and 1995–2011—Continued

Grade 8

Average score 1995	Average score 2007	Average score 2011	Benchmarking education systems
	556 ▲	567 ▲	*Massachusetts-USA*[2,4]
544 ▲	539 ▲	553 ▲	*Minnesota-USA*[4]
550 ▲		546 ▲	*Alberta-CAN*[2]
496 ▼	526	521	*Ontario-CAN*[2]
510	507 ▼	520	*Quebec-CAN*
	489 ▼	485 ▼	*Dubai-UAE*



▲ Score is higher than U.S. score.
▼ Score is lower than U.S. score.
Change from 2007 to 2011.
Change from 1995 to 2011.
Rounds to zero.
*p<.05. Change in average scores is significant.
[1]The change in average score is calculated by subtracting the 2007 or 1995 estimate, respectively, from the 2011 estimate using unrounded numbers.
[2]National Defined Population covers 90 to 95 percent of National Target Population for 2011 (see appendix A).
[3]Nearly satisfied guidelines for sample participation rates after replacement schools were included for 2011.
[4]National Target Population does not include all of the International Target Population for 2011 (see appendix A).
[5]Exclusion rates for Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available for 2011.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation in 2011 exceeds 15 percent, though it is less than 25 percent.
NOTE: Education systems are ordered by 2011 average scores. Italics indicate participants identified and counted in this report as an education system and not as a separate country. All education systems met international sampling and other guidelines in 2011, except as noted. Data are not shown for some education systems because comparable data from previous cycles are not available. Participants that did not administer TIMSS at the target grade are not shown; see the international report for their results. All U.S. state data are based on public school students only.
For 1995, Lithuania's National Target Population did not include all of the International Target Population; the Russian Federation and Lithuania had a National Defined Population that covered 90 to 95 percent of National Target Population; England-GBR, and Ontario-CAN had a National Defined Population that covered less than 90 percent of National Target Population (but at least 77 percent); the United States, England-GBR, and Minnesota-USA met guidelines for sample participation rates only after replacement schools were included. For 2007, Lithuania, Georgia, and Indonesia had National Target Populations that did not include all of the International Target Population; the United States, Massachusetts-USA, Minnesota-USA, and Ontario-CAN had National Defined Population that covered 90 to 95 percent of National Target Population; Hong Kong-CHN, England-GBR, and Minnesota-USA met guidelines for sample participation rates only after replacement schools were included; Dubai-UAE nearly satisfied guidelines for sample participation rates after replacement schools were included.
All average scores reported as higher or lower than the U.S. average score are different at the .05 level of statistical significance. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between averages for one education system may be significant, while a large difference for another education system may not be significant. Detail may not sum to totals because of rounding. The standard errors of the estimates are shown in table E-25 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 2007, and 2011.

Content domain scores in 2011

In addition to overall average science scores, TIMSS provides average scores by specific science topics called *content domains*. At grade 4, TIMSS tested student knowledge in three content domains: *life science*, *physical science*, and *Earth science*. At grade 8, TIMSS tested student knowledge in four content domains: *biology*, *chemistry*, *physics*, and *Earth science*.

At grade 4, the U.S. average score was higher than the TIMSS scale average of 500 in all three content domains in 2011 (table 28). In comparison with other education systems, U.S. 4th-graders performed better on average in *life science* than in the other two domains. That is, fewer education systems outperformed the United States in *life science* than in *physical science* or *Earth science*. U.S. 4th-graders were outperformed on average by their peers in 4

Table 28. Average science content domain scores of 4th-grade students, by education system: 2011

Education system	Life science	Physical science	Earth science
Singapore[1]	597 ▲	598 ▲	541
Finland	574 ▲	568 ▲	566 ▲
Korea, Rep. of	571 ▲	597 ▲	603 ▲
Russian Federation	556 ▲	548	552 ▲
Hungary	552	520 ▼	524 ▼
Czech Republic	550	519 ▼	537
United States[1]	**547**	**544**	**539**
Japan	540 ▼	589 ▲	551 ▲
Chinese Taipei-CHN	538 ▼	569 ▲	553 ▲
Netherlands[2]	537 ▼	526 ▼	525 ▼
Italy	535 ▼	509 ▼	523 ▼
Slovak Republic	534 ▼	527 ▼	535
Sweden	534 ▼	528 ▼	538
England-GBR	530 ▼	535 ▼	522 ▼
Denmark[1]	530 ▼	526 ▼	527 ▼
Austria	526 ▼	535 ▼	539
Germany	525 ▼	535 ▼	520 ▼
Croatia[1]	525 ▼	502 ▼	521 ▼
Hong Kong-CHN[1]	524 ▼	539	548 ▲
Slovenia	524 ▼	524 ▼	506 ▼
Portugal	520 ▼	517 ▼	531
Lithuania[1,3]	520 ▼	514 ▼	501 ▼
Northern Ireland-GBR[2]	519 ▼	520 ▼	507 ▼
Serbia[1]	518 ▼	523 ▼	497 ▼
Australia	516 ▼	514 ▼	520 ▼
Poland	514 ▼	495 ▼	496 ▼
Spain	513 ▼	497 ▼	499 ▼
Ireland	513 ▼	517 ▼	520 ▼
Belgium (Flemish)-BEL	510 ▼	507 ▼	505 ▼
Romania	504 ▼	508 ▼	502 ▼
Kazakhstan[1]	500 ▼	486 ▼	491 ▼
New Zealand	497 ▼	493 ▼	499 ▼
Norway[4]	496 ▼	482 ▼	506 ▼
Chile	490 ▼	471 ▼	475 ▼
Thailand	480 ▼	462 ▼	460 ▼
Georgia[3,5]	461 ▼	440 ▼	458 ▼
Turkey	460 ▼	466 ▼	456 ▼
Iran, Islamic Rep. of	449 ▼	453 ▼	457 ▼
Bahrain	444 ▼	453 ▼	445 ▼
Azerbaijan[1,5]	440 ▼	436 ▼	408 ▼
Malta	439 ▼	453 ▼	447 ▼
Armenia	424 ▼	399 ▼	398 ▼
United Arab Emirates	420 ▼	429 ▼	435 ▼
Saudi Arabia	415 ▼	439 ▼	432 ▼
Qatar[1]	383 ▼	397 ▼	401 ▼
Oman	370 ▼	370 ▼	371 ▼
Tunisia[6]	342 ▼	342 ▼	319 ▼
Kuwait[3,6]	323 ▼	348 ▼	352 ▼
Morocco[7]	245 ▼	256 ▼	208 ▼
Yemen[7]	172 ▼	198 ▼	186 ▼
Benchmarking education systems			
Florida-USA[3,8]	549	542	537
Alberta-CAN[1]	542	542	539
North Carolina-USA[1,3]	541	541	529
Ontario-CAN	535 ▼	528 ▼	514 ▼
Quebec-CAN	524 ▼	507 ▼	516 ▼
Dubai-UAE	455 ▼	460 ▼	469 ▼
Abu Dhabi-UAE	403 ▼	415 ▼	418 ▼

▲ Average score is higher than U.S. average score.
▼ Average score is lower than U.S. average score.
[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Met guidelines for sample participation rates only after replacement schools were included.
[3]National Target Population does not include all of the International Target Population (see appendix A).
[4]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[5]Exclusion rates for Azerbaijan and Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[7]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
[8]National Defined Population covers less than 90 percent, but at least 77 percent of National Target Population (see appendix A).
NOTE: Education systems are ordered by average score in *life science* domain. Italics indicate participants identified and counted in this report as an education system and not as a separate country. Participants that did not administer TIMSS at the target grade are not shown; see the international report for their results. All U.S. state data are based on public school students only. All average scores reported as higher or lower than U.S. average score are different at the .05 level of statistical significance. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one education system may be significant while a large difference between the United States and another education system may not be significant. The standard errors of the estimates are shown in table E-26 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

education systems in *life science*, 5 education systems in *physical science*, and 6 education systems in *Earth science*.

At grade 8, the U.S. average was also higher than the TIMSS scale average of 500 in all four content domains in 2011 (table 29). In comparison with other education systems, U.S. 8th-graders performed better on average in *biology* than in the other three domains. That is, fewer education systems had higher average scores than the United States in *biology* than in *chemistry*, *physics*, or *Earth science*. U.S. 8th-graders were outperformed on average by their peers in 9 education systems in *biology*, 10 education systems in *Earth science*, 10 education systems in *chemistry*, and 16 education systems in *physics*.

Table 29. Average science content domain scores of 8th-grade students, by education system: 2011

Education system	Biology	Chemistry	Physics	Earth science
Singapore[1]	594 ▲	590 ▲	602 ▲	566 ▲
Korea, Rep. of	561 ▲	551 ▲	577 ▲	548 ▲
Japan	561 ▲	560 ▲	558 ▲	548 ▲
Chinese Taipei-CHN	557 ▲	585 ▲	552 ▲	568 ▲
Finland	548 ▲	554 ▲	540 ▲	574 ▲
Russian Federation[1]	537	554 ▲	547 ▲	535
Hong Kong-CHN	535	526	539 ▲	539
England-GBR[2]	533	529	533 ▲	536
Slovenia	532	558 ▲	532 ▲	560 ▲
United States[1]	**530**	**520**	**513**	**533**
Australia	527	501 ▼	511	533
Israel[3]	523	514	514	504 ▼
Hungary	520 ▼	534 ▲	525 ▲	511 ▼
Lithuania[4]	517 ▼	517	503 ▼	517 ▼
New Zealand	514 ▼	501 ▼	509	523
Sweden	513 ▼	502 ▼	498 ▼	520 ▼
Italy	503 ▼	491 ▼	490 ▼	513 ▼
Ukraine	492 ▼	512	503 ▼	495 ▼
Norway	491 ▼	488 ▼	481 ▼	516 ▼
Turkey	484 ▼	477 ▼	494 ▼	468 ▼
Kazakhstan	483 ▼	508 ▼	489 ▼	472 ▼
Iran, Islamic Rep. of	466 ▼	469 ▼	483 ▼	477 ▼
United Arab Emirates	463 ▼	464 ▼	461 ▼	466 ▼
Chile	462 ▼	447 ▼	453 ▼	476 ▼
Thailand	460 ▼	436 ▼	430 ▼	466 ▼
Romania	458 ▼	469 ▼	456 ▼	470 ▼
Tunisia	449 ▼	434 ▼	436 ▼	414 ▼
Bahrain	449 ▼	448 ▼	457 ▼	451 ▼
Jordan	447 ▼	463 ▼	446 ▼	436 ▼
Georgia[4,5]	435 ▼	395 ▼	401 ▼	417 ▼
Saudi Arabia	430 ▼	428 ▼	437 ▼	441 ▼
Malaysia	427 ▼	426 ▼	435 ▼	401 ▼
Syrian Arab Republic	425 ▼	424 ▼	426 ▼	414 ▼
Armenia	420 ▼	452 ▼	441 ▼	421 ▼
Qatar	411 ▼	416 ▼	426 ▼	408 ▼
Indonesia	410 ▼	378 ▼	397 ▼	412 ▼
Oman	407 ▼	408 ▼	427 ▼	431 ▼
Palestinian Nat'l Auth.	407 ▼	432 ▼	432 ▼	406 ▼
Macedonia, Rep. of	400 ▼	416 ▼	398 ▼	403 ▼
Lebanon	395 ▼	435 ▼	405 ▼	365 ▼
Morocco	378 ▼	374 ▼	349 ▼	377 ▼
Ghana[6]	290 ▼	331 ▼	292 ▼	265 ▼
Benchmarking education systems				
Massachusetts-USA[1,4]	575 ▲	568 ▲	555 ▲	577 ▲
Minnesota-USA[4]	563 ▲	538 ▲	541 ▲	574 ▲
Alberta-CAN[1]	554 ▲	521	545 ▲	559 ▲
Colorado-USA[4]	551 ▲	528	530 ▲	555 ▲
North Carolina-USA[3,4]	541	531	510	540
Indiana-USA[1,4]	540	526	522	540
Connecticut-USA[1,4]	539	520	520	542
Ontario-CAN[1]	531	495 ▼	521 ▲	528
Florida-USA[1,4]	529	525	530 ▲	536
Quebec-CAN	525	515	502 ▼	536
California-USA[1,4]	500 ▼	503 ▼	487 ▼	499 ▼
Alabama-USA[4]	491 ▼	480 ▼	476 ▼	487 ▼
Dubai-UAE	485 ▼	487 ▼	482 ▼	487 ▼
Abu Dhabi-UAE	459 ▼	461 ▼	459 ▼	461 ▼

▲ Average score is higher than U.S. score.
▼ Average score is lower than U.S. score.
[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[3]National Defined Population covers less than 90 percent, but at least 77 percent of National Target Population (see appendix A).
[4]National Target Population does not include all of the International Target Population (see appendix A).
[5]Exclusion rates for Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
NOTE: Education systems are ordered by average score in *biology* domain. Italics indicate participants identified and counted in this report as an education system and not as a separate country. Participants that did not administer TIMSS at the target grade are not shown; see the international report for their results. All U.S. state data are based on public school students only. All average scores reported as higher or lower than U.S. average score are different at the .05 level of statistical significance. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one education system may be significant while a large difference between the United States and another education system may not be significant. The standard errors of the estimates are shown in table E-27 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Performance on the TIMSS international benchmarks

The TIMSS international benchmarks provide a way to understand how students' proficiency in science varies along the TIMSS scale (table 30). TIMSS defines four levels of student achievement: *Advanced*, *High*, *Intermediate*, and *Low*. The benchmarks can then be used to describe the kinds of skills and knowledge students at each score cutpoint would need to successfully answer the science items included in the assessment. The descriptions of the benchmarks differ between the two grade levels, as the scientific skills and knowledge needed to respond to the assessment items reflect the nature, difficulty, and emphasis at each grade.

In 2011, higher percentages of U.S. 4th-graders performed at or above each of the four TIMSS international benchmarks than the international medians (figure 11). For example, 15 percent of U.S. 4th-graders performed at or above the *Advanced* benchmark (625) compared to the international median of 5 percent. Students at the *Advanced* benchmark demonstrated an ability to apply their knowledge and understanding of scientific processes and relationships and show some knowledge of the process of scientific inquiry (see description in table 30).

The percentage of students performing at or above the *Advanced* international science benchmark was higher than in the United States in 3 education systems; was not different in 6 education systems; and was lower than the United States in 47 education systems.

Singapore, Korea, and Finland had a higher percentage of students performing at or above the *Advanced* international science benchmark than the United States at grade 4; and the Russian Federation, Chinese Taipei-CHN, Japan, Florida-USA, Hungary, and North Carolina-USA had a percentage that was not measurably different from the U.S. percentage.

Similar to their 4th-grade counterparts, higher percentages of U.S. 8th-graders performed at or above each of the four TIMSS international benchmarks than the international medians (figure 12). For example, 10 percent of U.S. 8th-graders performed at or above the *Advanced* benchmark (625) compared to the international median of 4 percent. Students at the *Advanced* benchmark demonstrated an ability to communicate an understanding of complex and abstract concepts in biology, chemistry, physics, and Earth science (see description in table 30).

The percentage of 8th-grade students performing at or above the *Advanced* international science benchmark was higher than in the United States in 12 education systems; was not different in 10 education systems; and was lower than in the United States in 33 education systems.

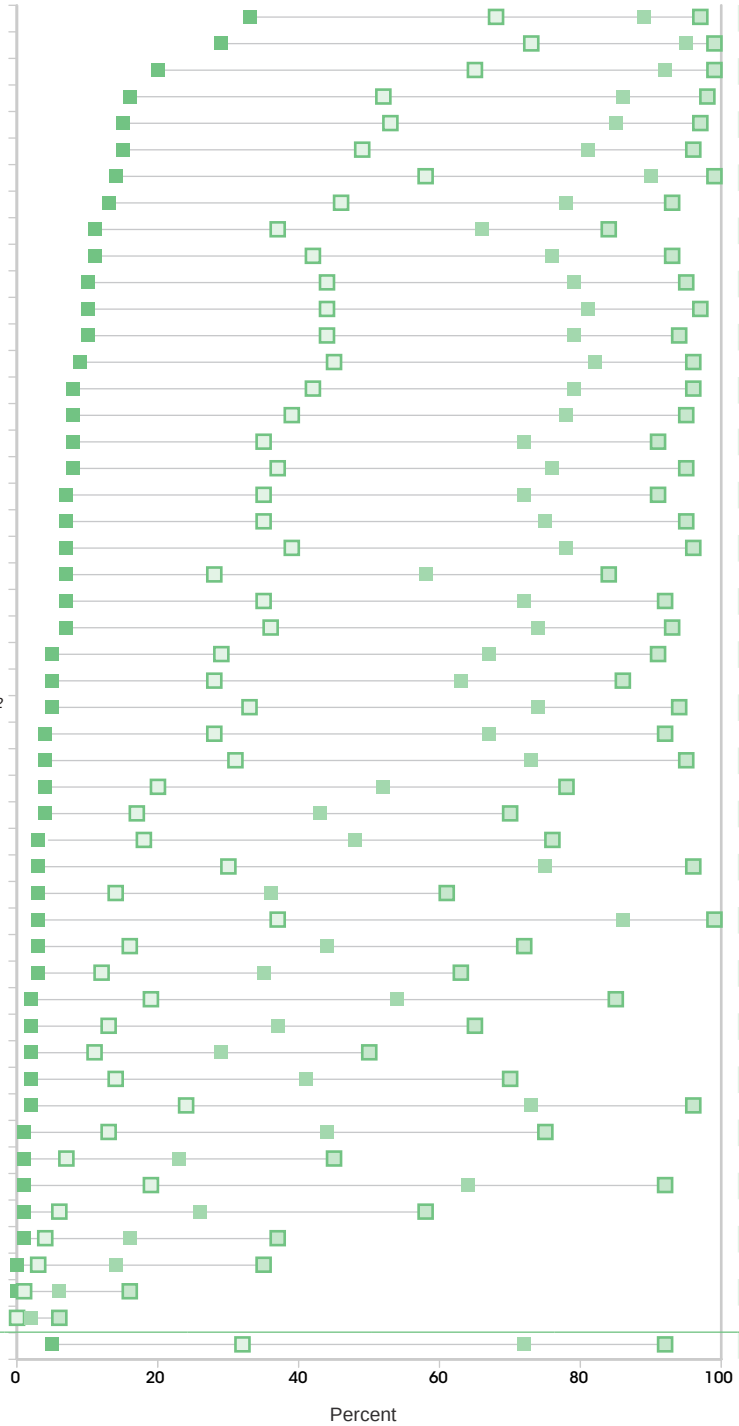
Singapore, Massachusetts-USA, Chinese Taipei-CHN, Korea, Japan, Minnesota-USA, Colorado-USA, Connecticut-USA, the Russian Federation, England-GBR, Slovenia, and Finland had a higher percentage of students performing at or above the *Advanced* international science benchmark than the United States at grade 8; and Florida-USA, North Carolina-USA, Alberta-CAN, Israel, Australia, Indiana-USA, Hong Kong-CHN, New Zealand, Hungary, and Turkey had a percentage that was not measurably different from the U.S. percentage.

Table 30. Description of TIMSS international science benchmarks, by grade: 2011

Benchmark (score cutpoint)	Grade 4
Advanced (625)	*Students apply knowledge and understanding of scientific processes and relationships and show some knowledge of the process of scientific inquiry.* Students communicate their understanding of characteristics and life processes of organisms, reproduction and development, ecosystems and organisms' interactions with the environment, and factors relating to human health. They demonstrate understanding of properties of light and relationships among physical properties of materials, apply and communicate their understanding of electricity and energy in practical contexts, and demonstrate an understanding of magnetic and gravitational forces and motion. Students communicate their understanding of the solar system and of Earth's structure, physical characteristics, resources, processes, cycles, and history. They have a beginning ability to interpret results in the context of a simple experiment, reason and draw conclusions from descriptions and diagrams, and evaluate and support an argument.
High (550)	*Students apply their knowledge and understanding of the sciences to explain phenomena in everyday and abstract contexts.* Students demonstrate some understanding of plant and animal structure, life processes, life cycles and reproduction. They also demonstrate some understanding of ecosystems and organisms' interactions with their environment, including understanding of human responses to outside conditions and activities. Students demonstrate understanding of some properties of matter, electricity and energy, and magnetic and gravitational forces and motion. They show some knowledge of the solar system, and of Earth's physical characteristics, processes, and resources. Students demonstrate elementary knowledge and skills related to scientific inquiry. They compare, contrast, and make simple inferences, and provide brief descriptive responses combining knowledge of science concepts with information from both everyday and abstract contexts.
Intermediate (475)	*Students have basic knowledge and understanding of practical situations in the sciences.* Students recognize some basic information related to characteristics of living things, their reproduction and life cycles, and their interactions with the environment, and show some understanding of human biology and health. They also show some knowledge of properties of matter and light, electricity and energy, and forces and motion. Students know some basic facts about the solar system and show an initial understanding of Earth's physical characteristics and resources. They demonstrate ability to interpret information in pictorial diagrams and apply factual knowledge to practical situations.
Low (400)	*Students have some elementary knowledge of life science and physical science.* Students demostrate a knowledge of some simple facts related to human health, ecosystems, and the behavioral and physical characteristics of animals. They also demonstrate some basic knowledge of energy and the physical properties of matter. Students interpret simple diagrams, complete simple tables, and provide short written responses to questions requiring factual information.
	Grade 8
Advanced (625)	*Students communicate an understanding of complex and abstract concepts in biology, chemistry, physics, and Earth science.* Students demonstrate some conceptual knowledge about cells and the characteristics, classification, and life processes of organisms. They communicate an understanding of the complexity of ecosystems and adaptations of organisms, and apply an understanding of life cycles and heredity. Students also communicate an understanding of the structure of matter and physical and chemical properties and changes and apply knowledge of forces, pressure, motion, sound, and light. They reason about electrical circuits and properties of magnets. Students apply knowledge and communicate understanding of the solar system and Earth's processes, structures, and physical features. They understand basic features of scientific investigation. They also combine information from several sources to solve problems and draw conclusions, and they provide written explanations to communicate scientific knowledge.
High (550)	*Students demonstrate understanding of concepts related to science cycles, systems, and principles.* They demonstrate understanding of aspects of human biology, and of the characteristics, classification, and life processes of organisms. Students communicate understanding of processes and relationships in ecosystems. They show an understanding of the classification and compositions of matter and chemical and physical properties and changes. They apply knowledge to situations related to light and sound and demonstrate basic knowledge of heat and temperature, forces and motion, and electrical circuits and magnets. Students demonstrate an understanding of the solar system and of Earth's processes, physical features, and resources. They demonstrate some scientific inquiry skills. They also combine and interpret information from various types of diagrams, contour maps, graphs, and tables; select relevant information, analyze, and draw conclusions; and provide short explanations conveying scientific knowledge.
Intermediate (475)	*Students recognize and apply their understanding of basic scientific knowledge in various contexts.* Students apply knowledge and communicate an understanding of human health, life cycles, adaptation, and heredity, and analyze information about ecosystems. They have some knowledge of chemistry in everyday life and elementary knowledge of properties of solutions and the concept of concentration. They are acquainted with some aspects of force, motion, and energy. They demonstrate an understanding of Earth's processes and physical features, including the water cycle and atmosphere. Students interpret information from tables, graphs, and pictorial diagrams and draw conclusions. They apply knowledge to practical situations and communicate their understanding through brief descriptive responses.
Low (400)	*Students can recognize some basic facts from the life and physical sciences.* They have some knowledge of the human body, and demonstrate some familiarity with physical phenomena. Students interpret simple pictorial diagrams, complete simple tables, and apply basic knowledge to practical situations.

NOTE: Score cutpoints for the international benchmarks are determined through scale anchoring. Scale anchoring involves selecting benchmarks (scale points) on the achievement scales to be described in terms of student performance, and then identifying items that students scoring at the anchor points can answer correctly. The score cutpoints are set at equal intervals along the achievement scales. The score cutpoints were selected to be as close as possible to the standard percentile cutpoints (i.e., 90th, 75th, 50th, and 25th percentiles). More information on the setting of the score cutpoints can be found in appendix A and in Martin et al. (2012).
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

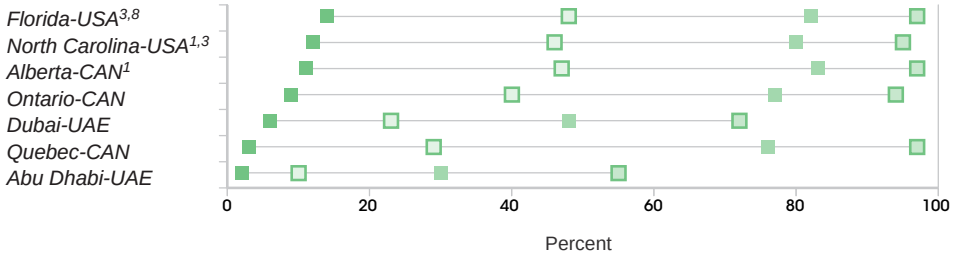
Figure 11. Percentage of 4th-grade students reaching the TIMSS international benchmarks in science, by education system: 2011



Education system	Percentage of students reaching each international benchmark			
	Advanced	High	Inter-mediate	Low
	(625)	(550)	(475)	(400)
Singapore[1]	33 *	68 *	89 *	97 *
Korea, Rep. of	29 *	73 *	95 *	99 *
Finland	20 *	65 *	92 *	99 *
Russian Federation	16	52	86 *	98 *
Chinese Taipei-CHN	15	53 *	85 *	97 *
United States[1]	**15**	**49**	**81**	**96**
Japan	14	58 *	90 *	99 *
Hungary	13	46	78 *	93 *
Romania	11 *	37 *	66 *	84 *
England-GBR	11 *	42 *	76 *	93 *
Sweden	10 *	44 *	79	95
Czech Republic	10 *	44 *	81	97
Slovak Republic	10 *	44 *	79	94
Hong Kong-CHN[1]	9 *	45	82	96
Austria	8 *	42 *	79	96
Denmark[1]	8 *	39 *	78 *	95
Serbia[1]	8 *	35 *	72 *	91 *
Italy	8 *	37 *	76 *	95
Australia	7 *	35 *	72 *	91 *
Portugal	7 *	35 *	75 *	95
Germany	7 *	39 *	78 *	96
Kazakhstan[1]	7 *	28 *	58 *	84 *
Ireland	7 *	35 *	72 *	92 *
Slovenia	7 *	36 *	74 *	93 *
Poland	5 *	29 *	67 *	91 *
New Zealand	5 *	28 *	63 *	86 *
Northern Ireland-GBR[2]	5 *	33 *	74 *	94
Spain	4 *	28 *	67 *	92 *
Lithuania[1,3]	4 *	31 *	73 *	95
Thailand	4 *	20 *	52 *	78 *
Bahrain	4 *	17 *	43 *	70 *
Turkey	3 *	18 *	48 *	76 *
Croatia[1]	3 *	30 *	75 *	96
United Arab Emirates	3 *	14 *	36 *	61 *
Netherlands[2]	3 *	37 *	86 *	99 *
Iran, Islamic Rep. of	3 *	16 *	44 *	72 *
Saudi Arabia	3 *	12 *	35 *	63 *
Chile	2 *	19 *	54 *	85 *
Azerbaijan[1,4]	2 *	13 *	37 *	65 *
Qatar[1]	2 *	11 *	29 *	50 *
Malta	2 *	14 *	41 *	70 *
Belgium (Flemish)-BEL	2 *	24 *	73 *	96
Georgia[3,4]	1 *	13 *	44 *	75 *
Oman	1 *	7 *	23 *	45 *
Norway[5]	1 *	19 *	64 *	92 *
Armenia	1 *	6 *	26 *	58 *
Kuwait[3,6]	1 *	4 *	16 *	37 *
Morocco[7]	# *	3 *	14 *	35 *
Tunisia[6]	# *	1 *	6 *	16 *
Yemen[7]	# *	# *	2 *	6 *
International Median	**5 ***	**32 ***	**72 ***	**92 ***

See notes at end of table.

Figure 11. Percentage of 4th-grade students reaching the TIMSS international benchmarks in science, by education system: 2011—Continued

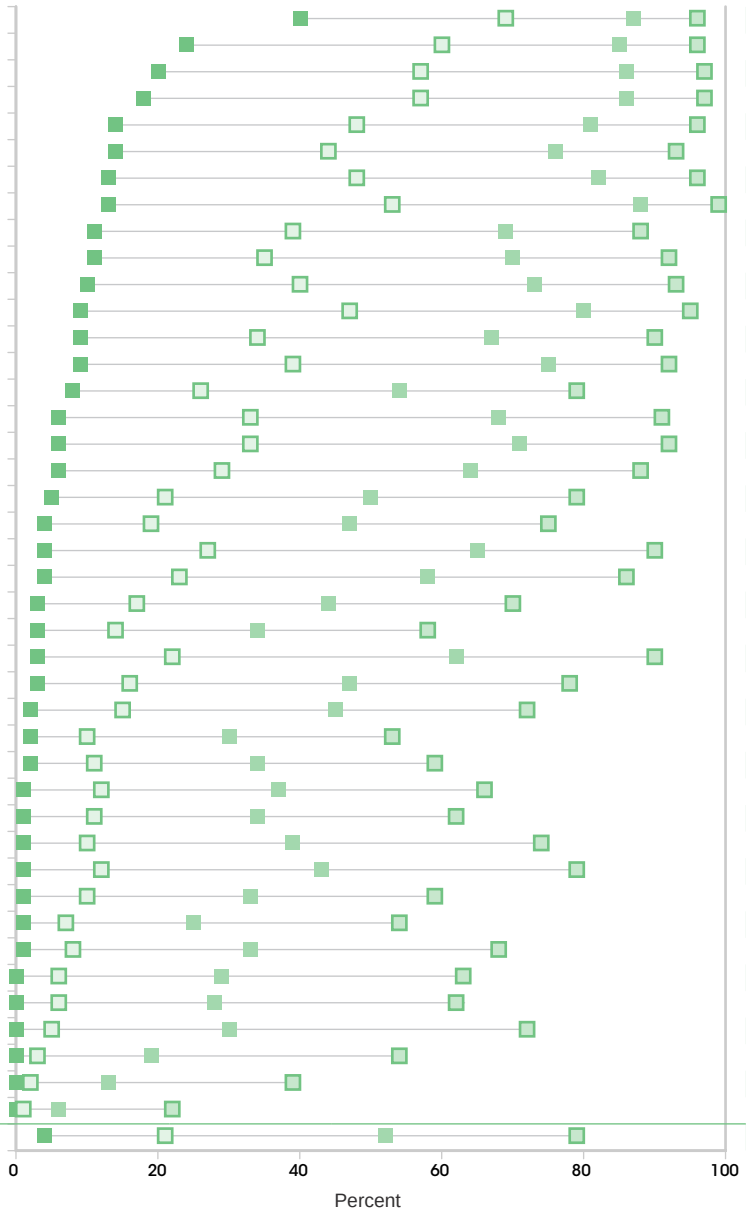


Benchmarking education systems	Percentage of students reaching each international benchmark			
	Advanced	High	Inter-mediate	Low
	(625)	(550)	(475)	(400)
Florida-USA[3,8]	14	48	82	97 *
North Carolina-USA[1,3]	12	46	80	95
Alberta-CAN[1]	11 *	47	83	97 *
Ontario-CAN	9 *	40 *	77 *	94
Dubai-UAE	6 *	23 *	48 *	72 *
Quebec-CAN	3 *	29 *	76 *	97 *
Abu Dhabi-UAE	2 *	10 *	30 *	55 *

Advanced benchmark
High benchmark
Intermediate benchmark
Low benchmark

Rounds to zero.
*p<.05. Percentage is significantly different from the U.S. percentage at the same benchmark.
[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Met guidelines for sample participation rates only after replacement schools were included.
[3]National Target Population does not include all of the International Target Population (see appendix A).
[4]Exclusion rates for Azerbaijan and Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[5]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[7]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
[8]National Defined Population covers less than 90 percent, but at least 77 percent of National Target Population (see appendix A).
NOTE: Education systems are ordered by percentage at *Advanced* international benchmark. Italics indicate participants identified and counted in this report as an education system and not as a separate country. The TIMSS international median represents all participating TIMSS education systems, including the United States, shown in the main part of the figure; benchmarking education systems are not included in the median. Participants that did not administer TIMSS at the target grade are not shown; see the international report for their results. All U.S. state data are based on public school students only. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one education system may be significant while a large difference between the United States and another education system may not be significant. The standard errors of the estimates are shown in table E-28 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Figure 12. Percentage of 8th-grade students reaching the TIMSS international benchmarks in science, by education system: 2011

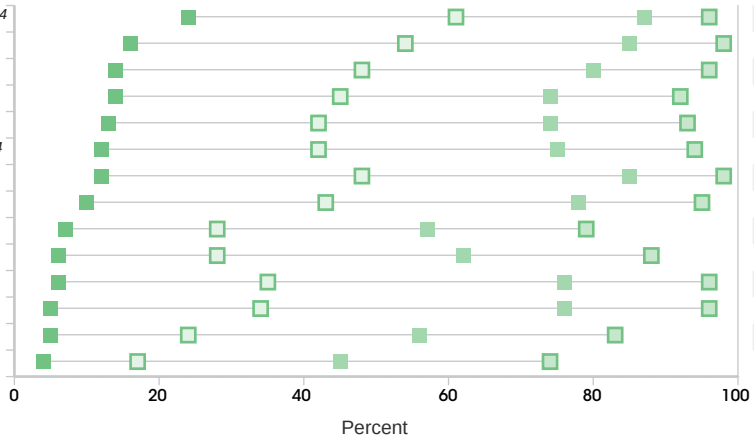


Education system	Advanced (625)	High (550)	Intermediate (475)	Low (400)
Singapore[1]	40 *	69 *	87 *	96 *
Chinese Taipei-CHN	24 *	60 *	85 *	96 *
Korea, Rep. of	20 *	57 *	86 *	97 *
Japan	18 *	57 *	86 *	97 *
Russian Federation[1]	14 *	48 *	81 *	96 *
England-GBR[2]	14 *	44	76	93
Slovenia	13 *	48 *	82 *	96 *
Finland	13 *	53 *	88 *	99 *
Israel[3]	11	39	69 *	88 *
Australia	11	35	70	92
United States[1]	**10**	**40**	**73**	**93**
Hong Kong-CHN	9	47 *	80 *	95
New Zealand	9	34 *	67 *	90 *
Hungary	9	39	75	92
Turkey	8	26 *	54 *	79 *
Sweden	6 *	33 *	68 *	91 *
Lithuania[4]	6 *	33 *	71	92
Ukraine	6 *	29 *	64 *	88 *
Iran, Islamic Rep. of	5 *	21 *	50 *	79 *
United Arab Emirates	4 *	19 *	47 *	75 *
Italy	4 *	27 *	65 *	90 *
Kazakhstan	4 *	23 *	58 *	86 *
Bahrain	3 *	17 *	44 *	70 *
Qatar	3 *	14 *	34 *	58 *
Norway	3 *	22 *	62 *	90 *
Romania	3 *	16 *	47 *	78 *
Jordan	2 *	15 *	45 *	72 *
Macedonia, Rep. of	2 *	10 *	30 *	53 *
Oman	2 *	11 *	34 *	59 *
Armenia	1 *	12 *	37 *	66 *
Malaysia	1 *	11 *	34 *	62 *
Thailand	1 *	10 *	39 *	74 *
Chile	1 *	12 *	43 *	79 *
Palestinian Nat'l Auth.	1 *	10 *	33 *	59 *
Lebanon	1 *	7 *	25 *	54 *
Saudi Arabia	1 *	8 *	33 *	68 *
Syrian Arab Republic	# *	6 *	29 *	63 *
Georgia[4,5]	# *	6 *	28 *	62 *
Tunisia	# *	5 *	30 *	72 *
Indonesia	# *	3 *	19 *	54 *
Morocco	# *	2 *	13 *	39 *
Ghana[6]	# *	1 *	6 *	22 *
International Median	**4 ***	**21 ***	**52 ***	**79 ***

Percentage of students reaching each international benchmark

See notes at end of table.

Figure 12. Percentage of 8th-grade students reaching the TIMSS international benchmarks in science, by education system: 2011—Continued



Benchmarking education systems	Percentage of students reaching each international benchmark			
	Advanced	High	Intermediate	Low
	(625)	(550)	(475)	(400)
Massachusetts-USA[1,4]	24 *	61 *	87 *	96 *
Minnesota-USA[4]	16 *	54 *	85 *	98 *
Colorado-USA[4]	14 *	48 *	80 *	96 *
Connecticut-USA[1,4]	14 *	45	74	92
Florida-USA[1,4]	13	42	74	93
North Carolina-USA[3,4]	12	42	75	94
Alberta-CAN[1]	12	48 *	85 *	98 *
Indiana-USA[1,4]	10	43	78	95 *
Dubai-UAE	7 *	28 *	57 *	79 *
California-USA[1,4]	6 *	28 *	62 *	88 *
Ontario-CAN[1]	6 *	35 *	76	96 *
Quebec-CAN	5 *	34 *	76	96 *
Alabama-USA[4]	5 *	24 *	56 *	83 *
Abu Dhabi-UAE	4 *	17 *	45 *	74 *

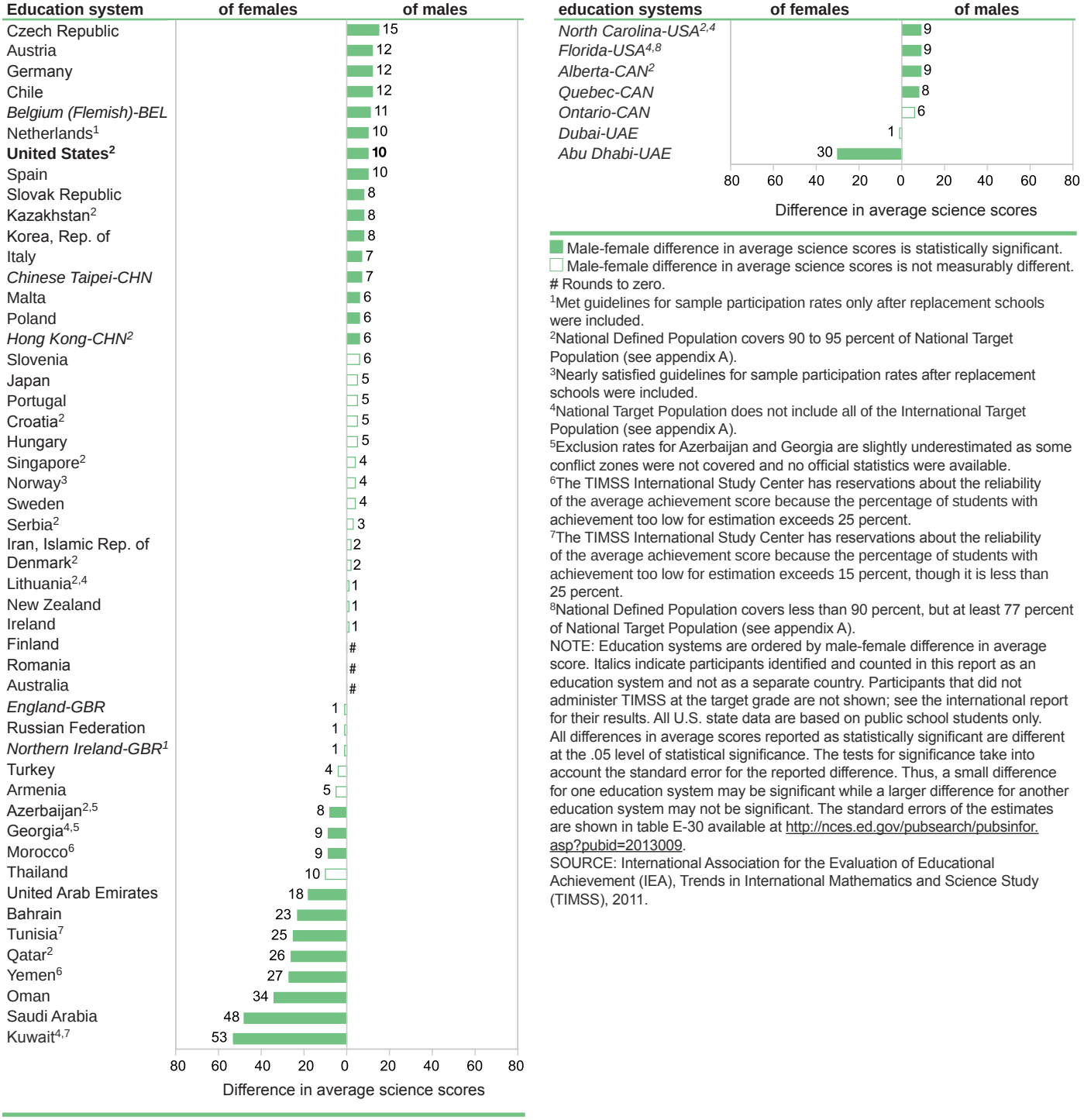
Advanced benchmark
High benchmark
Intermediate benchmark
Low benchmark

Rounds to zero.
*p<.05. Percentage is significantly different from the U.S. percentage at the same benchmark.
[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[3]National Defined Population covers less than 90 percent, but at least 77 percent of National Target Population (see appendix A).
[4]National Target Population does not include all of the International Target Population (see appendix A).
[5]Exclusion rates for Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
NOTE: Education systems are ordered by percentage at *Advanced* international benchmark. Italics indicate participants identified and counted in this report as an education system and not as a separate country. The TIMSS international median represents all participating TIMSS education systems, including the United States, shown in the main part of the figure; benchmarking education systems are not included in the median. Participants that did not administer TIMSS at the target grade are not shown; see the international report for their results. All U.S. state data are based on public school students only. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one education system may be significant while a large difference between the United States and another education system may not be significant. The standard errors of the estimates are shown in table E-29 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Average scores of male and female students

In 2011, the U.S. average score in science at grade 4 was 10 points higher for males than for females (figure 13). Among all 57 education systems that participated in TIMSS at grade 4, there were 32 education systems that showed a measurable difference in the average science scores of males and females: 20 in favor of males (including both participating U.S. states) and 12 in favor of females. The difference in average scores between males and females ranged from 53 score points in Kuwait (in favor of females) to 15 score points in the Czech

Figure 13. Difference in average science scores of 4th-grade students, by sex and education system: 2011

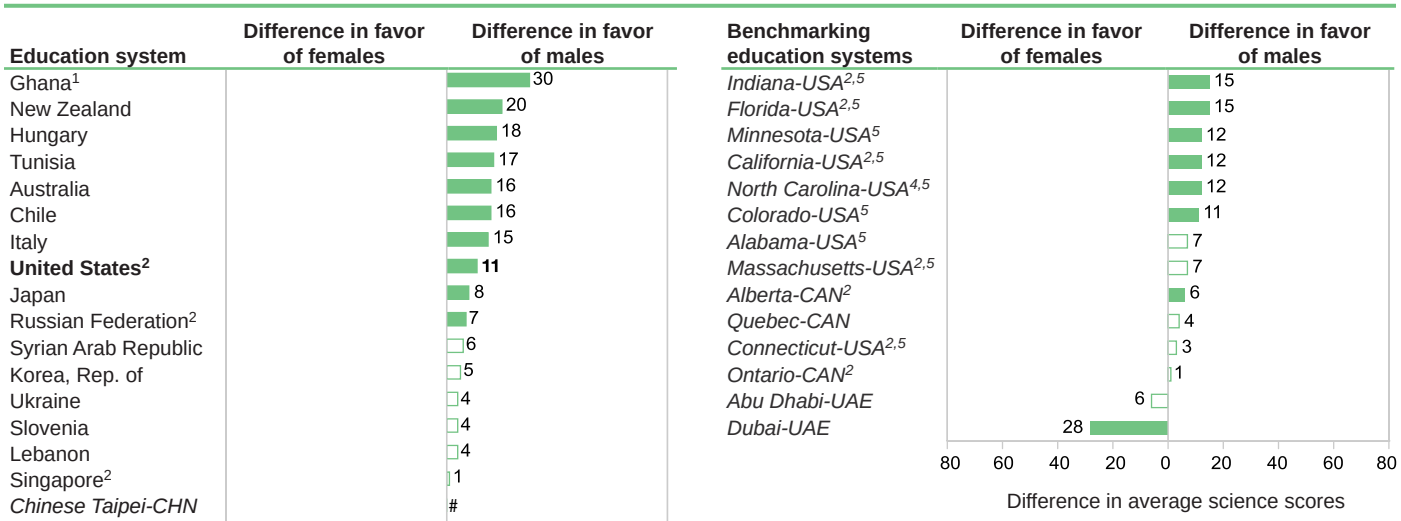


Male-female difference in average science scores is statistically significant.
Male-female difference in average science scores is not measurably different.
Rounds to zero.
[1]Met guidelines for sample participation rates only after replacement schools were included.
[2]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[3]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[4]National Target Population does not include all of the International Target Population (see appendix A).
[5]Exclusion rates for Azerbaijan and Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
[7]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[8]National Defined Population covers less than 90 percent, but at least 77 percent of National Target Population (see appendix A).
NOTE: Education systems are ordered by male-female difference in average score. Italics indicate participants identified and counted in this report as an education system and not as a separate country. Participants that did not administer TIMSS at the target grade are not shown; see the international report for their results. All U.S. state data are based on public school students only. All differences in average scores reported as statistically significant are different at the .05 level of statistical significance. The tests for significance take into account the standard error for the reported difference. Thus, a small difference for one education system may be significant while a larger difference for another education system may not be significant. The standard errors of the estimates are shown in table E-30 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Republic (in favor of males). In 25 education systems, there was no measurable difference between the average science scores of males and females.

At grade 8, the U.S. average score in science was 11 points higher for males than for females (figure 14). Among all 56 education systems that participated in TIMSS at grade 8, there were 33 education systems that showed a significant difference in the average science scores of males and females: 17 in favor of males (including all the participating U.S. states except Alabama, Connecticut, and Massachusetts, which had no measurable difference in average scores between males

Figure 14. Difference in average science scores of 8th-grade students, by sex and education system: 2011



Male-female difference in average science scores is statistically significant.
Male-female difference in average science scores is not measurably different.
Rounds to zero.
[1]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[2]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[3]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[4]National Defined Population covers less than 90 percent, but at least 77 percent of National Target Population (see appendix A).
[5]National Target Population does not include all of the International Target Population (see appendix A).
[6]Exclusion rates for Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
NOTE: Education systems are ordered by male-female difference in average score. Italics indicate participants identified and counted in this report as an education system and not as a separate country. Participants that did not administer TIMSS at the target grade are not shown; see the international report for their results. All U.S. state data are based on public school students only. All differences in average scores reported as statistically significant are different at the .05 level of statistical significance. The tests for significance take into account the standard error for the reported difference. Thus, a small difference for one education system may be significant while a larger difference for another education system may not be significant. The standard errors of the estimates are shown in table E-31 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

and females) and 16 in favor of females. The difference in average scores between males and females ranged from 78 score points in Oman (in favor of females) to 30 score points in Ghana (in favor of males). In 23 education systems, there was no measurable difference between the average science scores of males and females.

Performance within the United States

In 2011, TIMSS was administered to enough students and in enough schools in the United States to provide separate average science scores for students by race/ethnicity and schools serving varying percentages of low-income students as measured by the percentage of students eligible for free or reduced-price lunch.

In addition, TIMSS was administered to enough students and in enough schools in nine U.S. states to provide each of the states its own separate TIMSS results for public school students at grade 8 and, in two of the states, at grade 4 as well. These state science results are reported at the end of this section.

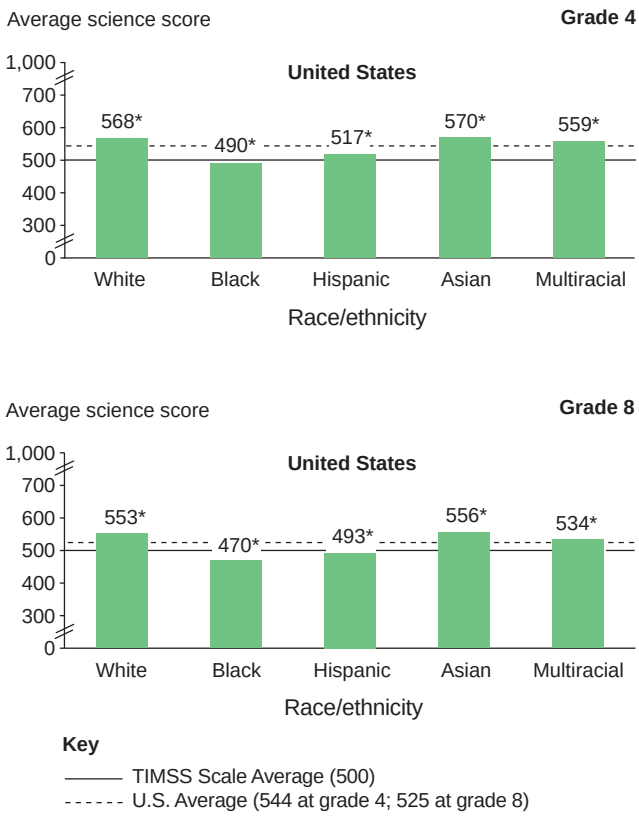
As mentioned in the introduction (and explained in detail in appendix A), separate state public school samples were drawn, at grade 4, for Florida and North Carolina and, at grade 8, for Alabama, California, Colorado, Connecticut, Florida, Indiana, Massachusetts, Minnesota, and North Carolina. Some of these states chose to participate as benchmarking participants in order to compare their performance internationally, and others were invited to participate in TIMSS by the National Assessment of Educational Progress (NAEP), which is conducting a study to link TIMSS and NAEP (as explained in appendix A). The states invited to participate at grade 8 were selected based on state enrollment size and willingness to participate, as well as on their general NAEP performance (above or below the national average on NAEP), their previous experience in benchmarking to TIMSS, and their regional distribution.

Average scores of students of different races and ethnicities

In 2011, the average science scores for U.S. White, Asian, Hispanic, and multiracial 4th-graders were higher than the TIMSS scale average, but for U.S. Black 4th-graders the average was lower (figure 15). In comparison with the U.S. national average, U.S. White, Asian, and multiracial 4th-graders scored higher, on average, while U.S. Black and Hispanic 4th-graders scored lower, on average.

At grade 8, the average science scores for U.S. White, Asian, and multiracial students were higher than both the TIMSS scale average and the U.S. national average. However, U.S. Black and Hispanic 8th-graders scored lower, on average, than the TIMSS scale average and U.S. national average.

Figure 15. Average science scores of U.S. 4th- and 8th-grade students, by race/ethnicity: 2011



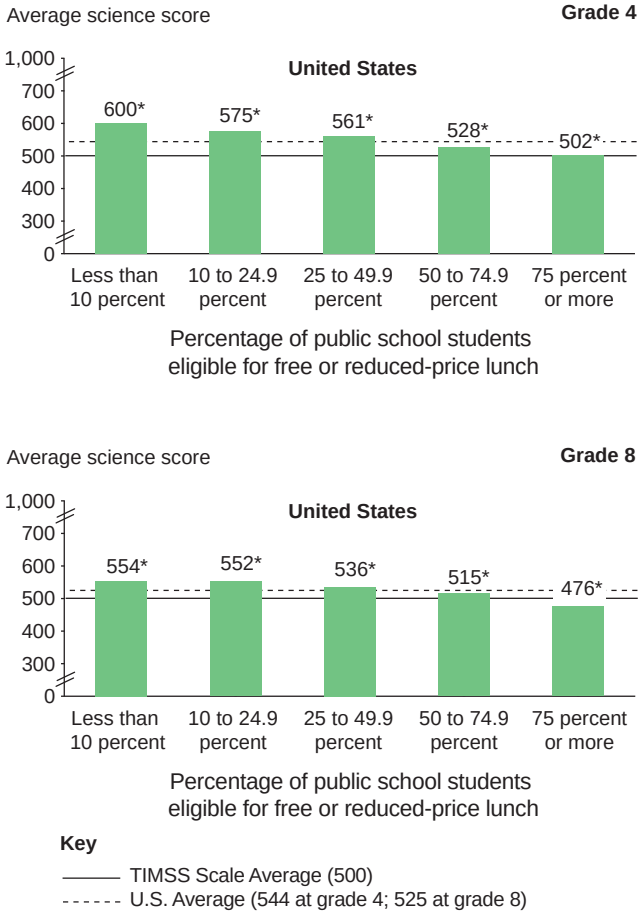
*p<.05. Difference between score and U.S. average score is significant.
NOTE: Reporting standards were not met for American Indian/Alaska Native and Native Hawaiian/Other Pacific Islander. Black includes African American, Hispanic includes Latino, and Asian includes Pacific Islander and Native Hawaiian. Racial categories exclude Hispanic origin. Students who identified themselves as being of Hispanic origin were classified as Hispanic, regardless of their race. Although data for some race/ethnicities are not shown separately because the reporting standards were not met, they are included in the U.S. and state totals shown throughout the report. See appendix A in this report for more information. The standard errors of the estimates are shown in table E-32 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Average scores of students attending public schools of various poverty levels

In 2011, the average science score of U.S. 4th-graders in the highest poverty public schools (at least 75 percent of students eligible for free or reduced-price lunch) was not measurably different from the TIMSS scale average. The average scores of 4th-graders in each of the other categories of school poverty were higher than the TIMSS scale average (figure 16). Fourth-graders in the highest poverty public schools, as well as those in public schools with at least 50 percent but less than 75 percent of students eligible for free or reduced-price lunch, had average scores below the U.S. national average, while those in public schools with lower proportions of low-income students scored higher, on average, than the U.S. national average.

At grade 8, students in the highest poverty public schools had a lower average score than the TIMSS scale average (476 vs. 500), while students in public schools with less than 75 percent of students eligible for free or reduced-price lunch had a higher score, on average, than the TIMSS scale average. Eighth-graders in public schools with less than 50 percent of students eligible for free or reduced-price lunch scored, on average, above the U.S. national average, while students in public schools with 50 percent or more of students eligible scored, on average, below.

Figure 16. Average science scores of U.S. 4th- and 8th-grade students, by percentage of public school students eligible for free or reduced-price lunch: 2011



*p<.05. Difference between score and U.S. average score is significant.
NOTE: Analyses are limited to public schools only, based on school reports of the percentage of students in public school eligible for the federal free or reduced-price lunch program. The standard errors of the estimates are shown in table E-33 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

TIMSS 2011 results for Alabama

Science – Grade 8

- Public school students' average score was 485 at grade 8.
- The percentages of Alabama 8th-graders reaching the *Advanced*, *High*, and *Intermediate* international science benchmarks were not measurably different from the international medians (figure 12). However, Alabama 8th-graders performed above the international median at the *Low* benchmark (83 percent vs. 79 percent).
- Female students in Alabama scored below the TIMSS scale average in science (table 32).
- White students' average scores were higher than the TIMSS scale average. Asian and multiracial students' average scores were not measurably different from the TIMSS scale average. However, Black and Hispanic students scored lower, on average, that the TIMSS scale average (table 32).
- Students in public schools with less than 10 percent of students eligible for free or reduced-price lunch scored higher, on average, than the TIMSS scale average, while those in schools with 75 percent or more of students eligible for free or reduced-price lunch scored lower, on average, than the TIMSS scale average.

Table 31. Average science scores of 8th-grade students in Alabama public schools compared with other participating education systems: 2011

Grade 8	
Education systems higher than Alabama	
Singapore	*North Carolina-USA*
Massachusetts-USA	*Florida-USA*
Chinese Taipei-CHN	United States
Korea, Rep. of	Hungary
Japan	*Ontario-CAN*
Minnesota-USA	*Quebec-CAN*
Finland	Australia
Alberta-CAN	Israel
Slovenia	Lithuania
Russian Federation	New Zealand
Colorado-USA	Sweden
Hong Kong-CHN	Italy
England-GBR	Ukraine
Indiana-USA	
Connecticut-USA	
Education systems not measurably different from Alabama	
California-USA	*Dubai-UAE*
Norway	Turkey
Kazakhstan	Iran, Islamic Rep. of
Education systems lower than Alabama	
Romania	Syrian Arab Republic
United Arab Emirates	*Palestinian Nat'l Auth.*
Chile	Georgia
Abu Dhabi-UAE	Oman
Bahrain	Qatar
Thailand	Macedonia, Rep. of
Jordan	Lebanon
Tunisia	Indonesia
Armenia	Morocco
Saudi Arabia	Ghana
Malaysia	

NOTE: Italics indicate participants identified and counted in this report as an education system and not as a separate country.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Table 32. Average science scores in grade 8 for selected student groups in public schools in Alabama: 2011

	Science
Reporting groups	**Grade 8**
TIMSS scale average	**500**
U.S. average	**525 ***
Alabama average	**485 ***
Sex	
Female	482 *
Male	489
Race/ethnicity	
White	519 *
Black	435 *
Hispanic	470 *
Asian	493
Multiracial	511
Percentage of public school students eligible for free or reduced-price lunch	
Less than 10 percent	557 *
10 to 24.9 percent	521
25 to 49.9 percent	504
50 to 74.9 percent	492
75 percent or more	441 *

*$p<.05$. Difference between score and TIMSS scale average is significant.
NOTE: Black includes African American, Hispanic includes Latino, and Asian includes Pacific Islander and Native Hawaiian. Racial categories exclude Hispanic origin. Not all race/ethnicity categories are shown, but they are all included in the U.S. and state totals shown throughout the report. The standard errors of the estimates are shown in table E-34 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

TIMSS 2011 results for California

Science – Grade 8

- Public school students' average score was 499 at grade 8.
- Higher percentages of California 8th-graders performed at or above each of the four TIMSS international benchmarks than the international medians. For example, 6 percent of 8th-graders in California performed at or above the *Advanced* benchmark (625) compared to the international median of 4 percent at grade 8 (figure 12).
- In California males outperformed females by 12 score points, on average, in science (figure 14).
- White, Asian, and multiracial students' average scores were higher than the TIMSS scale average, while Black and Hispanic students scored lower, on average, than the TIMSS scale average.
- Students in public schools with less than 50 percent of students eligible for free or reduced-price lunch scored higher, on average, than the TIMSS scale average, while students in public schools with 75 percent or more of students eligible for free or reduced-price lunch scored lower, on average, than the TIMSS scale average.

Table 33. Average science scores of 8th-grade students in California public schools compared with other participating education systems: 2011

Grade 8	
Education systems higher than California	
Singapore	*Indiana-USA*
Massachusetts-USA	*Connecticut-USA*
Chinese Taipei-CHN	*North Carolina-USA*
Korea, Rep. of	*Florida-USA*
Japan	United States
Minnesota-USA	Hungary
Finland	*Ontario-CAN*
Alberta-CAN	*Quebec-CAN*
Slovenia	Australia
Russian Federation	Israel
Colorado-USA	Lithuania
Hong Kong-CHN	New Zealand
England-GBR	Sweden
Education systems not measurably different from California	
Italy	Norway
Ukraine	Kazakhstan
	Alabama-USA
Education systems lower than California	
Dubai-UAE	Saudi Arabia
Turkey	Malaysia
Iran, Islamic Rep. of	Syrian Arab Republic
Romania	*Palestinian Nat'l Auth.*
United Arab Emirates	Georgia
Chile	Oman
Abu Dhabi-UAE	Qatar
Bahrain	Macedonia, Rep. of
Thailand	Lebanon
Jordan	Indonesia
Tunisia	Morocco
Armenia	Ghana

NOTE: Italics indicate participants identified and counted in this report as an education system and not as a separate country.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Table 34. Average science scores in grade 8 for selected student groups in public schools in California: 2011

Reporting groups	Science Grade 8
TIMSS scale average	**500**
U.S. average	**525 ***
California average	**499**
Sex	
Female	493
Male	504
Race/ethnicity	
White	546 *
Black	460 *
Hispanic	475 *
Asian	542 *
Multiracial	529 *
Percentage of public school students eligible for free or reduced-price lunch	
Less than 10 percent	547 *
10 to 24.9 percent	542 *
25 to 49.9 percent	539 *
50 to 74.9 percent	493
75 percent or more	457 *

*p<.05. Difference between score and TIMSS scale average is significant.
NOTE: Black includes African American, Hispanic includes Latino, and Asian includes Pacific Islander and Native Hawaiian. Racial categories exclude Hispanic origin. Not all race/ethnicity categories are shown, but they are all included in the U.S. and state totals shown throughout the report. The standard errors of the estimates are shown in table E-35 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

TIMSS 2011 results for Colorado

Science – Grade 8

- Public school students' average score was 542 at grade 8.
- Higher percentages of Colorado 8th-graders performed at or above each of the four TIMSS international benchmarks than the international medians. For example, 14 percent of 8th-graders in Colorado performed at or above the *Advanced* benchmark (625) compared to the international median of 4 percent at grade 8 (figure 12).
- In Colorado males outperformed females by 11 score points, on average, in science (figure 14). Male and female students in Colorado scored higher, on average, in science than the TIMSS scale average (table 36).
- White, Asian, and multiracial students' average scores were higher than the TIMSS scale average. Black and Hispanic students' average scores were not measurably different (table 36).
- Students in public schools with at least 10 percent but less than 50 percent of students eligible for free or reduced-price lunch scored higher, on average, than the TIMSS scale average. Average scores of students in public schools in other categories of eligibility for free or reduced-price lunch were not measurably different from the TIMSS scale average, including those in schools where less than 10 percent were eligible.

Table 35. Average science scores of 8th-grade students in Colorado public schools compared with other participating education systems: 2011

Grade 8	
Education systems higher than Colorado	
Singapore	Korea, Rep. of
Massachusetts-USA	Japan
Chinese Taipei-CHN	Finland
Education systems not measurably different from Colorado	
Minnesota-USA	*England-GBR*
Alberta-CAN	*Indiana-USA*
Slovenia	*Connecticut-USA*
Russian Federation	*North Carolina-USA*
Hong Kong-CHN	*Florida-USA*
Education systems lower than Colorado	
United States	Chile
Hungary	*Abu Dhabi-UAE*
Ontario-CAN	Bahrain
Quebec-CAN	Thailand
Australia	Jordan
Israel	Tunisia
Lithuania	Armenia
New Zealand	Saudi Arabia
Sweden	Malaysia
Italy	Syrian Arab Republic
Ukraine	*Palestinian Nat'l Auth.*
California-USA	Georgia
Norway	Oman
Kazakhstan	Qatar
Alabama-USA	Macedonia, Rep. of
Dubai-UAE	Lebanon
Turkey	Indonesia
Iran, Islamic Rep. of	Morocco
Romania	Ghana
United Arab Emirates	

NOTE: Italics indicate participants identified and counted in this report as an education system and not as a separate country.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Table 36. Average science scores in grade 8 for selected student groups in public schools in Colorado: 2011

	Science
Reporting groups	**Grade 8**
TIMSS scale average	**500**
U.S. average	**525 ***
Colorado average	**542 ***
Sex	
Female	537 *
Male	548 *
Race/ethnicity	
White	572 *
Black	507
Hispanic	499
Asian	549 *
Multiracial	552 *
Percentage of public school students eligible for free or reduced-price lunch	
Less than 10 percent	534
10 to 24.9 percent	568 *
25 to 49.9 percent	560 *
50 to 74.9 percent	514
75 percent or more	486

*p<.05. Difference between score and TIMSS scale average is significant.
NOTE: Black includes African American, Hispanic includes Latino, and Asian includes Pacific Islander and Native Hawaiian. Racial categories exclude Hispanic origin. Not all race/ethnicity categories are shown, but they are all included in the U.S. and state totals shown throughout the report. The standard errors of the estimates are shown in table E-36 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

TIMSS 2011 results for Connecticut

Science – Grade 8

- Public school students' average score was 532 at grade 8.
- Higher percentages of Connecticut 8th-graders performed at or above each of the four TIMSS international benchmarks than the international medians. For example, 14 percent of 8th-graders in Connecticut performed at or above the *Advanced* benchmark (625) compared to the international median of 4 percent at grade 8 (figure 12).
- Male and female students in Connecticut scored higher, on average, in science than the TIMSS scale average.
- White, Asian, and multiracial students' average scores were higher than the TIMSS scale average, while Black and Hispanic students scored lower, on average, than the TIMSS scale average (table 38).
- Students in public schools with less than 25 percent of students eligible for free or reduced-price lunch scored higher, on average, than the TIMSS scale average, while students in schools with 50 percent or more of students eligible for free or reduced-price lunch scored lower, on average, than the TIMSS scale average.

Table 37. Average science scores of 8th-grade students in Connecticut public schools compared with other participating education systems: 2011

Grade 8	
Education systems higher than Connecticut	
Singapore	*Minnesota-USA*
Massachusetts-USA	Finland
Chinese Taipei-CHN	*Alberta-CAN*
Korea, Rep. of	Slovenia
Japan	
Education systems not measurably different from Connecticut	
Russian Federation	*North Carolina-USA*
Colorado-USA	Florida-USA
Hong Kong-CHN	United States
England-GBR	Hungary
Indiana-USA	Australia
Education systems lower than Connecticut	
Ontario-CAN	*Abu Dhabi-UAE*
Quebec-CAN	Bahrain
Israel	Thailand
Lithuania	Jordan
New Zealand	Tunisia
Sweden	Armenia
Italy	Saudi Arabia
Ukraine	Malaysia
California-USA	Syrian Arab Republic
Norway	*Palestinian Nat'l Auth.*
Kazakhstan	Georgia
Alabama-USA	Oman
Dubai-UAE	Qatar
Turkey	Macedonia, Rep. of
Iran, Islamic Rep. of	Lebanon
Romania	Indonesia
United Arab Emirates	Morocco
Chile	Ghana

NOTE: Italics indicate participants identified and counted in this report as an education system and not as a separate country.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Table 38. Average science scores in grade 8 for selected student groups in public schools in Connecticut: 2011

Reporting groups	Science Grade 8
TIMSS scale average	**500**
U.S. average	**525 ***
Connecticut average	**532 ***
Sex	
Female	530 *
Male	533 *
Race/ethnicity	
White	562 *
Black	459 *
Hispanic	474 *
Asian	565 *
Multiracial	543 *
Percentage of public school students eligible for free or reduced-price lunch	
Less than 10 percent	581 *
10 to 24.9 percent	549 *
25 to 49.9 percent	509
50 to 74.9 percent	471
75 percent or more	420 *

*p<.05. Difference between score and TIMSS scale average is significant.
NOTE Black includes African American, Hispanic includes Latino, and Asian includes Pacific Islander and Native Hawaiian. Racial categories exclude Hispanic origin. Not all race/ethnicity categories are shown, but they are all included in the U.S. and state totals shown throughout the report. The standard errors of the estimates are shown in table E-37 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

TIMSS 2011 results for Florida

Science – Grades 4 and 8

- Public school students' average science score was 545 at grade 4 and 530 at grade 8.
- Higher percentages of Florida 4th- and 8th-graders performed at or above each of the four TIMSS international benchmarks than the international medians. For example, 14 percent of 4th-graders and 13 percent of 8th-graders in Florida performed at or above the *Advanced* benchmark (625) compared to the international median of 5 percent at grade 4 and 4 percent at grade 8 (figures 11 and 12).
- Males outperformed females by 9 score points on average in science at grade 4 (figure 13) and by 15 score points at grade 8 (figure 14). In both grade 4 and grade 8, male and female students in Florida scored higher, on average, in science than the TIMSS scale average (table 40).

Table 39. Average science scores of 4th- and 8th-grade students in Florida public schools compared with other participating education systems: 2011

Grade 4	
Education systems higher than Florida	
Korea, Rep. of	Finland
Singapore	Japan
Education systems not measurably different from Florida	
Russian Federation	*North Carolina-USA*
Chinese Taipei-CHN	Czech Republic
United States	*Hong Kong-CHN*
Alberta-CAN	
Education systems lower than Florida	
Hungary	New Zealand
Sweden	Kazakhstan
Slovak Republic	Norway
Austria	Chile
Netherlands	Thailand
England-GBR	Turkey
Denmark	*Dubai-UAE*
Germany	Georgia
Ontario-CAN	Iran, Islamic Rep. of
Italy	Bahrain
Portugal	Malta
Slovenia	Azerbaijan
Northern Ireland-GBR	Saudi Arabia
Quebec-CAN	United Arab Emirates
Ireland	Armenia
Croatia	*Abu Dhabi-UAE*
Australia	Qatar
Serbia	Oman
Lithuania	Kuwait
Belgium (Flemish)-BEL	Tunisia
Romania	Morocco
Spain	Yemen
Poland	

Grade 8	
Education systems higher than Florida	
Singapore	Japan
Massachusetts-USA	*Minnesota-USA*
Chinese Taipei-CHN	Finland
Korea, Rep. of	*Alberta-CAN*
Education systems not measurably different from Florida	
Slovenia	*North Carolina-USA*
Russian Federation	United States
Colorado-USA	Hungary
Hong Kong-CHN	*Ontario-CAN*
England-GBR	*Quebec-CAN*
Indiana-USA	Australia
Connecticut-USA	Israel
Education systems lower than Florida	
Lithuania	Syrian Arab Republic
New Zealand	*Palestinian Nat'l Auth.*
Sweden	Georgia
Italy	Oman
Ukraine	Qatar
California-USA	Macedonia, Rep. of
Norway	Lebanon
Kazakhstan	Indonesia
Alabama-USA	Morocco
Dubai-UAE	Ghana
Turkey	
Iran, Islamic Rep. of	
Romania	
United Arab Emirates	
Chile	
Abu Dhabi-UAE	
Bahrain	
Thailand	
Jordan	
Tunisia	
Armenia	
Saudi Arabia	
Malaysia	

NOTE: Italics indicate participants identified and counted in this report as an education system and not as a separate country.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

- At grade 4 and grade 8, White, Hispanic, Asian, and multiracial students' average scores were higher than the TIMSS scale average, while Black students' average scores were not measurably different from the TIMSS scale average (table 40).
- Students at grade 4 scored higher, on average, than the TIMSS scale average regardless of the level of poverty within public schools. At grade 8, students in public schools with at least 10 percent but less than 75 percent of students eligible for free or reduced-price lunch scored, on average, higher than the TIMSS scale average, while the average score for students in public schools with 75 percent or more of students eligible for free or reduced-price lunch was not measurably different from the TIMSS scale average.

Table 40. Average science scores in grade 4 and 8 for selected student groups in public schools in Florida: 2011

	Science	
Reporting groups	**Grade 4**	**Grade 8**
TIMSS scale average	**500**	**500**
U.S. average	**544 ***	**525 ***
Florida average	**545 ***	**530 ***
Sex		
Female	540 *	522 *
Male	549 *	537 *
Race/ethnicity		
White	575 *	560 *
Black	504	485
Hispanic	531 *	523 *
Asian	593 *	600 *
Multiracial	577 *	524 *
Percentage of public school students eligible for free or reduced-price lunch		
Less than 10 percent	613 *	‡
10 to 24.9 percent	599 *	566 *
25 to 49.9 percent	556 *	550 *
50 to 74.9 percent	541 *	530 *
75 percent or more	517 *	498

‡ Reporting standards not met.
*p<.05. Difference between score and TIMSS scale average is significant.
NOTE: Black includes African American, Hispanic includes Latino, and Asian includes Pacific Islander and Native Hawaiian. Racial categories exclude Hispanic origin. Not all race/ethnicity categories are shown, but they are all included in the U.S. and state totals shown throughout the report. The standard errors of the estimates are shown in table E-38 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

TIMSS 2011 results for Indiana

Science – Grade 8

- Public school students' average score was 533 at grade 8.
- Higher percentages of Indiana 8th-graders performed at or above each of the four TIMSS international benchmarks than the international medians. For example, 10 percent of 8th-graders in Indiana performed at or above the *Advanced* benchmark (625) compared to the international median of 4 percent at grade 8 (figure 12).
- In Indiana, males outperformed females by 15 score points on average in science at grade 8 (figure 14). Male and female students in Indiana scored higher, on average, in science than the TIMSS scale average (table 42).
- White students' average scores were higher than the TIMSS scale average, while Black and Asian students scored lower, on average, than the TIMSS scale average. Hispanic and multiracial students' average scores were not measurably different from the TIMSS scale average.
- Students in public schools with at least 10 percent but less than 75 percent of students eligible for free or reduced-price lunch scored higher, on average, than the TIMSS scale average, while the average score of students in schools with 75 percent or more of students eligible for free or reduced-price lunch was not measurably different from the TIMSS scale average.

Table 41. Average science scores of 8th-grade students in Indiana public schools compared with other participating education systems: 2011

Grade 8	
Education systems higher than Indiana	
Singapore	Japan
Massachusetts-USA	*Minnesota-USA*
Chinese Taipei-CHN	Finland
Korea, Rep. of	*Alberta-CAN*
Education systems not measurably different from Indiana	
Slovenia	*Connecticut-USA*
Russian Federation	*North Carolina-USA*
Colorado-USA	*Florida-USA*
Hong Kong-CHN	United States
England-GBR	Hungary
Education systems lower than Indiana	
Ontario-CAN	*Abu Dhabi-UAE*
Quebec-CAN	Bahrain
Australia	Thailand
Israel	Jordan
Lithuania	Tunisia
New Zealand	Armenia
Sweden	Saudi Arabia
Italy	Malaysia
Ukraine	Syrian Arab Republic
California-USA	*Palestinian Nat'l Auth.*
Norway	Georgia
Kazakhstan	Oman
Alabama-USA	Qatar
Dubai-UAE	Macedonia, Rep. of
Turkey	Lebanon
Iran, Islamic Rep. of	Indonesia
Romania	Morocco
United Arab Emirates	Ghana
Chile	

NOTE: Italics indicate participants identified and counted in this report as an education system and not as a separate country.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Table 42. Average science scores in grade 8 for selected student groups in public schools in Indiana: 2011

	Science
Reporting groups	**Grade 8**
TIMSS scale average	**500**
U.S. average	**525 ***
Indiana average	**533 ***
Sex	
Female	526 *
Male	541 *
Race/ethnicity	
White	546 *
Black	460 *
Hispanic	499
Asian	492 *
Multiracial	534
Percentage of public school students eligible for free or reduced-price lunch	
Less than 10 percent	‡
10 to 24.9 percent	563 *
25 to 49.9 percent	540 *
50 to 74.9 percent	519 *
75 percent or more	476

‡ Reporting standards not met
*$p<.05$. Difference between score and TIMSS scale average is significant.
NOTE: Black includes African American, Hispanic includes Latino, and Asian includes Pacific Islander and Native Hawaiian. Racial categories exclude Hispanic origin. Not all race/ethnicity categories are shown, but they are all included in the U.S. and state totals shown throughout the report. The standard errors of the estimates are shown in table E-39 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

TIMSS 2011 results for Massachusetts

Science – Grade 8

- Public school students' average score was 567 at grade 8.
- Higher percentages of Massachusetts 8th-graders performed at or above each of the four TIMSS international benchmarks than the international medians. For example, 24 percent of 8th-graders in Massachusetts performed at or above the *Advanced* benchmark (625) compared to the international median of 4 percent at grade 8 (figure 12).

Table 43. Average science scores of 8th-grade students in Massachusetts public schools compared with other participating education systems: 2011

Grade 8	
Education systems higher than Massachusetts	
Singapore	
Education systems not measurably different from Massachusetts	
Chinese Taipei-CHN	Japan
Korea, Rep. of	*Minnesota-USA*
Education systems lower than Massachusetts	
Finland	*Alabama-USA*
Alberta-CAN	*Dubai-UAE*
Slovenia	Turkey
Russian Federation	Iran, Islamic Rep. of
Colorado-USA	Romania
Hong Kong-CHN	United Arab Emirates
England-GBR	Chile
Indiana-USA	*Abu Dhabi-UAE*
Connecticut-USA	Bahrain
North Carolina-USA	Thailand
Florida-USA	Jordan
United States	Tunisia
Hungary	Armenia
Ontario-CAN	Saudi Arabia
Quebec-CAN	Malaysia
Australia	Syrian Arab Republic
Israel	*Palestinian Nat'l Auth.*
Lithuania	Georgia
New Zealand	Oman
Sweden	Qatar
Italy	Macedonia, Rep. of
Ukraine	Lebanon
California-USA	Indonesia
Norway	Morocco
Kazakhstan	Ghana

NOTE: Italics indicate participants identified and counted in this report as an education system and not as a separate country.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

- Male and female students in Massachusetts scored higher, on average, in science than the TIMSS scale average (table 44).
- White, Asian, and multiracial students' average scores were higher than the TIMSS scale average, while Black and Hispanic students' average scores were not measurably different from the TIMSS scale average (table 44).
- Students in public schools with 75 percent or more of students eligible for free or reduced-price lunch had average scores not measurably different from the TIMSS scale average. All other groups scored, on average, above the TIMSS scale average.

Table 44. Average science scores in grade 8 for selected student groups in public schools in Massachusetts: 2011

	Science
Reporting groups	**Grade 8**
TIMSS scale average	**500**
U.S. average	**525 ***
Massachusetts average	**567 ***
Sex	
Female	564 *
Male	570 *
Race/ethnicity	
White	587 *
Black	514
Hispanic	494
Asian	576 *
Multiracial	576 *
Percentage of public school students eligible for free or reduced-price lunch	
Less than 10 percent	594 *
10 to 24.9 percent	589 *
25 to 49.9 percent	553 *
50 to 74.9 percent	550 *
75 percent or more	477

*p<.05. Difference between score and TIMSS scale average is significant.
NOTE: Black includes African American, Hispanic includes Latino, and Asian includes Pacific Islander and Native Hawaiian. Racial categories exclude Hispanic origin. Not all race/ethnicity categories are shown, but they are all included in the U.S. and state totals shown throughout the report. The standard errors of the estimates are shown in table E-40 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

TIMSS 2011 results for Minnesota

Science – Grade 8

- Public school students' average score was 553 at grade 8.
- Higher percentages of Minnesota 8th-graders performed at or above each of the four TIMSS international benchmarks than the international medians. For example, 16 percent of 8th-graders in Minnesota performed at or above the *Advanced* benchmark (625) compared to the international median of 4 percent at grade 8 (figure 12).
- In Minnesota, males outperformed females by 12 score points on average in science at grade 8 (figure 14). Male and female students scored higher, on average, in science than the TIMSS scale average (table 46).
- White and multiracial students' average scores were higher than the TIMSS scale average while Black, Hispanic, and Asian students' average scores were not measurably different from the TIMSS scale average.
- Students in public schools with 75 percent or more of students eligible for free or reduced-price lunch had average scores lower than the TIMSS scale average. All other groups scored, on average, above the TIMSS scale average.

Table 45. Average science scores of 8th-grade students in Minnesota public schools compared with other participating education systems: 2011

Grade 8	
Education systems higher than Minnesota	
Singapore	*Chinese Taipei-CHN*
Education systems not measurably different from Minnesota	
Massachusetts-USA	*Alberta-CAN*
Korea, Rep. of	Slovenia
Japan	Russian Federation
Finland	*Colorado-USA*
Education systems lower than Minnesota	
Hong Kong-CHN	Iran, Islamic Rep. of
England-GBR	Romania
Indiana-USA	United Arab Emirates
Connecticut-USA	Chile
North Carolina-USA	*Abu Dhabi-UAE*
Florida-USA	Bahrain
United States	Thailand
Hungary	Jordan
Ontario-CAN	Tunisia
Quebec-CAN	Armenia
Australia	Saudi Arabia
Israel	Malaysia
Lithuania	Syrian Arab Republic
New Zealand	*Palestinian Nat'l Auth.*
Sweden	Georgia
Italy	Oman
Ukraine	Qatar
California-USA	Macedonia, Rep. of
Norway	Lebanon
Kazakhstan	Indonesia
Alabama-USA	Morocco
Dubai-UAE	Ghana
Turkey	

NOTE: Italics indicate participants identified and counted in this report as an education system and not as a separate country.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Table 46. Average science scores in grade 8 for selected student groups in public schools in Minnesota: 2011

Reporting groups	Science Grade 8
TIMSS scale average	**500**
U.S. average	**525 ***
Minnesota average	**553 ***
Sex	
Female	548 *
Male	559 *
Race/ethnicity	
White	570 *
Black	489
Hispanic	512
Asian	511
Multiracial	537 *
Percentage of public school students eligible for free or reduced-price lunch	
Less than 10 percent	578 *
10 to 24.9 percent	570 *
25 to 49.9 percent	547 *
50 to 74.9 percent	555 *
75 percent or more	458 *

*$p<.05$. Difference between score and TIMSS scale average is significant.
NOTE: Black includes African American, Hispanic includes Latino, and Asian includes Pacific Islander and Native Hawaiian. Racial categories exclude Hispanic origin. Not all race/ethnicity categories are shown, but they are all included in the U.S. and state totals shown throughout the report. The standard errors of the estimates are shown in table E-41 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

TIMSS 2011 results for North Carolina

Science – Grades 4 and 8

- Public school students' average science score was 538 at grade 4 and 532 at grade 8.
- Higher percentages of North Carolina 4th- and 8th-graders performed at or above each of the four TIMSS international benchmarks than the international medians. For example, 12 percent of 4th-graders and 12 percent of 8th-graders in North Carolina performed at or above the *Advanced* benchmark (625) compared to the international median of 5 percent at grade 4 and 4 percent at grade 8 (figures 11 and 12).
- Males outperformed females by 9 score points on average in science at grade 4 and by 12 score points at grade 8 (figures 13 and 14). At both grade 4 and grade 8, male and female students in North Carolina scored higher, on average, in science than the TIMSS scale average (table 48).
- At grade 4, White, Hispanic, Asian, and multiracial students scored, on average, above the TIMSS scale average. Black students' average scores were not measurably different from the TIMSS scale average.

Continued on next page

Table 47. Average science scores of 4th– and 8th-grade students in North Carolina public schools compared with other participating education systems: 2011

Grade 4		Grade 8	
Education systems higher than North Carolina		**Education systems higher than North Carolina**	
Korea, Rep. of	Japan	Singapore	Japan
Singapore	Russian Federation	*Massachusetts-USA*	*Minnesota-USA*
Finland	*Chinese Taipei-CHN*	*Chinese Taipei-CHN*	Finland
		Korea, Rep. of	*Alberta-CAN*
Education systems not measurably different from North Carolina		**Education systems not measurably different from North Carolina**	
Florida-USA	Slovak Republic	Slovenia	*Florida-USA*
United States	Austria	Russian Federation	United States
Alberta-CAN	Netherlands	*Colorado-USA*	Hungary
Czech Republic	*England-GBR*	*Hong Kong-CHN*	*Ontario-CAN*
Hong Kong-CHN	Denmark	*England-GBR*	*Quebec-CAN*
Hungary	Germany	*Indiana-USA*	Australia
Sweden	*Ontario-CAN*	*Connecticut-USA*	
Education systems lower than North Carolina		**Education systems lower than North Carolina**	
Italy	Thailand	Israel	Bahrain
Portugal	Turkey	Lithuania	Thailand
Slovenia	*Dubai-UAE*	New Zealand	Jordan
Northern Ireland-GBR	Georgia	Sweden	Tunisia
Quebec-CAN	Iran, Islamic Rep. of	Italy	Armenia
Ireland	Bahrain	Ukraine	Saudi Arabia
Croatia	Malta	*California-USA*	Malaysia
Australia	Azerbaijan	Norway	Syrian Arab Republic
Serbia	Saudi Arabia	Kazakhstan	*Palestinian Nat'l Auth.*
Lithuania	United Arab Emirates	*Alabama-USA*	Georgia
Belgium (Flemish)-BEL	Armenia	*Dubai-UAE*	Oman
Romania	*Abu Dhabi-UAE*	Turkey	Qatar
Spain	Qatar	Iran, Islamic Rep. of	Macedonia, Rep. of
Poland	Oman	Romania	Lebanon
New Zealand	Kuwait	United Arab Emirates	Indonesia
Kazakhstan	Tunisia	Chile	Morocco
Norway	Morocco	*Abu Dhabi-UAE*	Ghana
Chile	Yemen		

NOTE: Italics indicate participants identified and counted in this report as an education system and not as a separate country.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

- At grade 8, White and Asian students scored, on average, above the TIMSS scale average while Black students scored lower, on average. Hispanic and multiracial students' average scores were not measurably different from the TIMSS scale average.
- In general, at grade 4 students in public schools with less than 75 percent of students eligible for free or reduced-price lunch scored higher, on average, than the TIMSS scale average. Average scores among students in public schools with 75 percent or more of students eligible for free or reduced-price lunch were not measurably different from the TIMSS scale average. At grade 8, students in public schools with less than 50 percent of students eligible for free or reduced-price lunch scored higher, on average, than the TIMSS scale average, while average scores for students in schools with 50 percent or more students eligible for free or reduced-price lunch were not measurably different from the TIMSS scale average.

Table 48. Average science scores in grade 4 and 8 for selected student groups in public schools in North Carolina: 2011

	Science	
Reporting groups	**Grade 4**	**Grade 8**
TIMSS scale average	**500**	**500**
U.S. average	**544 ***	**525 ***
North Carolina average	**538 ***	**532 ***
Sex		
Female	534 *	526 *
Male	543 *	537 *
Race/ethnicity		
White	565 *	565 *
Black	492	481 *
Hispanic	519 *	502
Asian	590 *	577 *
Multiracial	553 *	513
Percentage of public school students eligible for free or reduced-price lunch		
Less than 10 percent	‡	595 *
10 to 24.9 percent	574 *	569 *
25 to 49.9 percent	555 *	538 *
50 to 74.9 percent	534 *	518
75 percent or more	498	504

‡ Reporting standards not met.
*p<.05. Difference between score and TIMSS scale average is significant.
NOTE: Black includes African American, Hispanic includes Latino, and Asian includes Pacific Islander and Native Hawaiian. Racial categories exclude Hispanic origin. Not all race/ethnicity categories are shown, but they are all included in the U.S. and state totals shown throughout the report. The standard errors of the estimates are shown in table E-42 available at http://nces.ed.gov/pubsearch/pubsinfor.asp?pubid=2013009.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

References

Beaton, A.E., and González, E. (1995). *The NAEP Primer*. Chestnut Hill, MA: Boston College.

Chowdhury, S., Chu, A., and Kaufman, S. (2001). Minimizing Overlap in NCES Surveys. *Proceedings of the Survey Methods Research Section, American Statistical Association*, pp. 174-179.

Ferraro, D., and Van de Kerckhove, W. (2006). *Trends in International Mathematics and Science Study (TIMSS) 2003: Nonresponse Bias Analysis* (NCES 2007-044). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Foy, P., Joncas, M., and Zuhlke, O. (2009).*TIMSS 2011 School Sampling Manual*. Unpublished manuscript, Chestnut Hill, MA: Boston College.

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., and Brenwald, S. (2008). *Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context* (NCES 2009-01 Revised). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

IEA Data Processing Center. (2010). *TIMSS 2011 Data Entry Manager Manual*. Hamburg, Germany: Author.

Martin, M.O., and Mullis, I.V.S. (Eds.). (2011). *TIMSS and PIRLS Methods and Procedures*. Retrieved from http://timssandpirls.bc.edu/methods/index.html.

Martin, M.O., Mullis, I.V.S., Foy, P., and Stanco, G.M. (2012). *TIMSS 2011 International Results in Science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Matheson, N., Salganik, L., Phelps, R., Perie, M., Alsalam, N., and Smith, T. (1996). *Education Indicators: An International Perspective* (NCES 96-003). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Mullis, I.V.S., Martin, M.O., and Foy, P. (2005). *IEA's TIMSS 2003 International Report on Achievement in the Mathematics Cognitive Domains: Findings From a Developmental Project*. Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Foy, P., and Arora, A. (2012). *TIMSS 2011 International Results in Mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., and Preuschoff, C. (2009). *TIMSS 2011 Assessment Frameworks*. Chestnut Hill, MA: Boston College.

National Center for Education Statistics. (2002). *NCES Statistical Standards* (NCES 2003-601). Institute of Education Sciences, U.S. Department of Education. Washington, DC: Author.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (1999). *Classifying Educational Programmes Manual for ISCED-97 Implementation in OECD Countries* (1999 Edition). Paris: Author. Retrieved April 9, 2008, from http://www.oecd.org/dataoecd/7/2/1962350.pdf.

Westat. (2007). *WesVar 5.0 User's Guide*. Rockville, MD: Author.

Page intentionally left blank

Appendix A: Technical Notes

Introduction

The Trends in International Mathematics and Science Study (TIMSS) is a cross-national comparative study of the performance and schooling contexts of 4th- and 8th-grade students in mathematics and science. In this fifth cycle of TIMSS, mathematics and science assessments and associated questionnaires were administered in 57 education systems (45 of which were countries) at the 4th-grade level and 56 education systems (38 of which were countries) at the 8th-grade level during fall 2010 (in the Southern hemisphere) and during spring 2011 (in the Northern hemisphere). TIMSS is coordinated by the International Association for the Evaluation of Educational Achievement (IEA), with governmental sponsors in each participating country or education system. In the United States, TIMSS is sponsored by the National Center for Education Statistics (NCES), in the Institute of Education Sciences of the U.S. Department of Education.

As part of the 2011 administration of TIMSS in the United States, NCES conducted a study to link the National Assessment of Educational Progress (NAEP), a national U.S. student assessment in mathematics and science, with TIMSS so that states can measure their performance against international benchmarks. This NAEP-TIMSS Linking Study uses 8th-grade mathematics and science data from NAEP to project state-level scores onto the TIMSS scale. The goal of the study is to predict 2011 TIMSS mathematics and science scores at 8th-grade based on state NAEP performance without incurring the costs associated with every state participating separately in TIMSS. Results of the study are forthcoming as part of a separate report and technical notes are included here only as the linking study related to technical aspects of the main study, the focus of this report.

This appendix provides an overview of technical aspects of TIMSS 2011, including

- International requirements for sampling design, data collection, and response rates;
- Sampling, data collection, and response rates in the United States and for all participants;
- Test development;
- Recruitment, test administration, and quality assurance;
- Scoring and scoring reliability;
- Weighting, scaling, and plausible values;
- International benchmarks;
- Data limitations;
- Description of background variables;
- Confidentiality and disclosure limitations; and
- Statistical procedures.

More detailed information can be found in the *TIMSS and PIRLS Methods and Procedures* (Martin and Mullis 2011).

International requirements for sampling, data collection, and response rates

In order to ensure comparability of the data across countries, the IEA provided detailed international requirements on the various aspects of data collection described here and implemented quality control procedures. Participating countries were obliged to follow these requirements. These requirements regarding the target populations, sampling design, sample size, exclusions, and defining participation rates are described below.

Target populations

In order to identify comparable populations of students to be sampled, the IEA defined the target populations as follows (Martin and Mullis 2011):

Fourth-grade student population. The international desired target population (also referred to as the International Target Population) is all students enrolled in the grade that represents 4 years of schooling, counting from the first year of the International Standard Classification of Education (ISCED) Level 1,[1] providing that the mean age at the time of testing is at least 9.5 years.

Eighth-grade student population. The international desired target population is all students enrolled in the grade that represents 8 years of schooling, counting from the first year of ISCED Level 1, providing that the mean age at the time of testing is at least 13.5 years.

Teacher population. The target population is all mathematics and science teachers linked to the selected students. Note that these teachers are not a representative sample of teachers within the country. Rather, they are the mathematics and science teachers who teach a representative sample of students in two grades within the country (grades 4 and 8 in the United States).

[1]The ISCED was developed by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) to facilitate the comparability of educational levels across countries. ISCED Level 1 begins with the first year of formal, academic learning (UNESCO 1999). In the United States, ISCED Level 1 begins at grade 1.

School population. The target population is all eligible schools[2] containing either one or more 4th-grade classrooms or one or more 8th-grade classrooms.

Although participating education systems were expected to include all students in the International Target Population, sometimes it was not feasible to include all of these students because of geographic or linguistic constraints specific to the country or territory. Thus, each participating education system had its own "national" desired target population (also referred to as the National Target Population), which was the International Target Population reduced by the exclusions of those sections of the population that were not possible to assess. Working from the National Target Population, each participating education system had to operationalize the definition of its population for sampling purposes: i.e., define their "national" defined target population (referred to as the National Defined Population). While each education system's National Defined Population ideally coincides with its National Target Population, in reality, there may be additional exclusions (e.g., of regions or school types) due to constraints of operationalizing the assessment.

Sampling design

It is not feasible to assess every 4th- and 8th-grade student in the United States. Thus, as is done in all participating countries and education systems, a representative sample of 4th- and 8th-grade students was selected. The sample design employed by the TIMSS 2011 assessment is generally referred to as a two-stage stratified cluster sample. The sampling units at each stage were defined as follows.

First-stage sampling units. In the first stage of sampling, sampling statisticians selected individual schools with a probability-proportionate-to-size (PPS) approach, which means that each school's probability of selection is proportional to the estimated number of students enrolled in the target grade. Prior to sampling, statisticians assigned schools in the sampling frame to a predetermined number of explicit or implicit strata. Then, sampling staff identified sample schools using a PPS systematic sampling method. Statisticians also identified substitute schools (schools to replace original sampled school that refused to participate). The original and substitute schools were identified simultaneously.

Second-stage sampling units. In the second stage of sampling, statisticians selected classrooms within sampled schools using sampling software provided by the International Study Center at Boston College. The software uses a sampling algorithm for selecting classes that standardized the class sampling across schools and assures uniformity in the class selection procedures across participants. The software identified at least one classroom from a list of eligible classrooms that sampling staff prepared for each target grade. In various countries and education systems, including the United States, more than one eligible classroom per target grade per school was selected when possible. All students in sampled classrooms were selected for assessment.

Sample size for the main survey

TIMSS guidelines call for a minimum of 150 schools to be sampled per grade, with a minimum of 4,000 students assessed per grade. The basic sample design of one classroom per target grade per school was designed to yield a total sample of approximately 4,500 students per population. Countries with small class sizes or less than 30 students per school were directed to consider sampling more schools, more classrooms per school, or both, to meet the minimum target of 4,000 tested students.

In the United States, a sample of 450 schools was drawn at 4th-grade and 600 schools at 8th-grade. These were larger sample sizes than used in previous administrations of TIMSS. The reason for a larger sample than in the past at 4th grade was that in 2011 both TIMSS (administered every 4 years) and another study sponsored by the IEA, the Progress in International Reading Literacy Study (PIRLS) (administered every 5 years at 4th grade), happened to coincide. Because the United States was participating in both studies, and because both studies required a 4th-grade sample of schools and students, the decision was made to draw a larger sample of schools and to request that both studies be administered in the same schools (where feasible), albeit to separate classroom samples of students.[3] Thus, TIMSS (4th grade) and PIRLS in the United States were administered in the same schools but to separately sampled classrooms of students.

The reason for a larger sample than in the past at 8th grade was that the NAEP-TIMSS Linking Study (described above) required that NAEP be administered in the same schools in which TIMSS was administered, with one intact mathematics classroom randomly assigned to TIMSS and another to the linking study. Given that in previous administrations of TIMSS, two 8th-grade classrooms were sampled per school to take TIMSS, this requirement led to a doubling of the previous TIMSS sample size (from 300 to 600 schools) so that the total number of sampled TIMSS classrooms remained the same.

[2]Some sampled schools may be considered ineligible for reasons noted in the "School exclusions" section below. All other schools are considered eligible.

[3]In some cases, sampled schools were unable to accommodate both studies due to small student enrollment in grade 4 or scheduling conflicts. Schools with at least two grade 4 classrooms were asked to participate in both studies, with one classroom being randomly assigned to TIMSS and the other to PIRLS. Up to two TIMSS classes and two PIRLS classes were selected in schools with sufficient student enrollment. In schools with only one grade 4 classroom, either the TIMSS or PIRLS assessment was randomly assigned, but not both. In no cases were the same students asked to complete both the TIMSS and PIRLS assessments at grade 4.

In addition to the national school samples at 4th and 8th grade, state samples were also drawn for two states at 4th grade and nine states at 8th grade. These additional school samples are described below.

Exclusions

The following discussion draws on the *TIMSS 2011 School Sampling Manual* (Foy, Joncas, and Zuhlke 2009). All schools and students excluded from the national defined target population are referred to as the excluded population. Exclusions could occur at the school level, with entire schools being excluded, or within schools, with specific students or entire classrooms excluded. TIMSS 2011 did not provide accommodations for students with disabilities or students who were unable to read or speak the language of the test. The IEA requirement with regard to exclusions is that they should not exceed more than 5 percent of the national desired target population (Foy, Joncas, and Zuhlke 2009). The specifications for school and student exclusions were applied equally to the U.S. national and state samples.

School exclusions. Education systems could exclude schools that

- are geographically inaccessible;
- are of extremely small size;
- offer a curriculum, or school structure, radically different from the mainstream educational system; or
- provide instruction only to students in the excluded categories defined under "within-school exclusions," such as schools for the blind.

Within-school exclusions. Education systems were asked to adopt the following international within-school exclusion rules to define excluded students:

- **Students with intellectual disabilities**—Students who, in the professional opinion of the school principal or other qualified staff members, are considered to have intellectual disabilities or who have been tested psychologically as such. This includes students who are emotionally or mentally unable to follow even the general instructions of the test. Students were not to be excluded solely because of poor academic performance or normal disciplinary problems.
- **Students with functional disabilities**—Students who are permanently physically disabled in such a way that they cannot perform in the TIMSS testing situation. Students with functional disabilities who are able to respond were to be included in the testing.
- **Non-English-language speakers**—Students who are unable to read or speak the language(s) of the test and would be unable to overcome the language barrier of the test. Typically, a student who had received less than one year of instruction in the language(s) of the test was to be excluded.

Defined participation rates

In order to minimize the potential for response biases, the IEA developed participation or response rate standards that apply to all participating education systems and govern both whether or not a participating education system's data are included in the TIMSS 2011 international dataset as well as the way in which national statistics are presented in the international reports. These standards were set using composites of response rates at the school, classroom, and student and teacher levels; moreover, response rates were calculated with and without the inclusion of substitute schools (selected to replace original sample schools refusing to participate).

The response rate standards determine how a participant's data will be reported in the international reports. These standards take the following two forms, distinguished primarily by whether or not meeting the school response rate of 85 percent requires the counting of substitute schools.

Category 1: Met requirements. Participants that meet all of the following conditions are considered to have fulfilled the IEA requirements: (a) a minimum school participation rate of 85 percent, based on original (sampled) schools only; (b) a minimum classroom participation rate of 95 percent, from both original and substitute schools; and (c) a minimum student participation rate of 85 percent, from both original and substitute schools.

Category 2: Met requirements after substitutes. In the case of participants not meeting the category 1 requirements, provided that at least 50 percent of schools in the original sample participate, a participating education system's data are considered acceptable if the following requirements are met: a minimum combined school, classroom, and student participation rate of 75 percent, based on the product of the participation rates described above. That is, the product of (a), (b), and (c), as defined in the category 1 standard, must be greater than or equal to 75 percent.

Participants satisfying the category 1 standard are included in the international tabular presentations without annotation. Those able to satisfy only the category 2 standard are included as well but are annotated to indicate their response rate status. The data from participants failing to meet either standard are presented separately in the international tabular presentations.

Sampling, data collection, and response rates in the United States and for all participants

The U.S. TIMSS national sample design

In the United States and most other participating countries and education systems, the target populations of students corresponded to the 4th and 8th grades. In sampling these populations, TIMSS used a two-stage stratified cluster sampling design (as explained above under "Sampling design").[4] The U.S. sampling frame was explicitly stratified by three categorical stratification variables: percentage of students eligible for free or reduced-price lunch, school control (public or private), and region of the country (Northeast, Central, West, Southeast).[5] The U.S. sampling frame was implicitly stratified (that is, sorted for sampling) by two categorical stratification variables: community type (four categories)[6] and minority status (i.e., above or below 15 percent of the student population).

The first stage made use of a systematic PPS technique to select schools for the original sample from a sampling frame based on the 2010 National Assessment of Educational Progress (NAEP) school sampling frame.[7] Data for public schools were taken from the Common Core of Data (CCD), and data for private schools were taken from the Private School Universe Survey (PSS). In addition, for each original school selected, the two neighboring schools in the sampling frame were designated as substitute schools. The first school following the original sample school was the first substitute and the first school preceding it was the second substitute. If an original school refused to participate, the first substitute was contacted. If that school also refused to participate, the second substitute was contacted. There were several constraints on the assignment of substitutes. One sampled school was not allowed to substitute for another, and a given school could not be assigned to substitute for more than one sampled school. Furthermore, substitutes were required to be in the same implicit stratum as the sampled school.

The second stage consisted of selecting intact mathematics classes within each participating school. Schools provided lists of 4th- or 8th-grade classrooms. Within schools, classrooms with fewer than 15 students were collapsed into *pseudo-classrooms*, so that each classroom in the school's classroom sampling frame had at least 20 students.[8] An equal probability sample of two classrooms[9] was identified from the classroom frame for the school. In schools where there was only one classroom, this classroom was selected with certainty. At the 4th-grade level, 16 pseudo-classrooms were created prior to classroom sampling and 8 of these were selected in the final 4th-grade classroom sample. At the 8th-grade level, 503 pseudo-classrooms were created, of which 64 were included in the final classroom sample. All students in sampled classrooms and pseudo-classrooms were selected for assessment.

In this way, the overall sample design for the United States results in an approximately self-weighting sample of students, with each 4th- or 8th-grade student having a roughly equal probability of selection. Note that in small schools, a higher proportion of the classes (and therefore of the students) is selected, but this higher rate of selecting students in small schools is offset by a lower selection rate of small schools, as schools are selected with probability proportional to size.

Additional sampling requirements for TIMSS 2011 in the United States: The NAEP-TIMSS Linking Study

The NAEP-TIMSS Linking Study administered NAEP and TIMSS booklets in both the NAEP and TIMSS administration windows under first NAEP and then TIMSS testing conditions. This was done to gather examinee-level correlations between NAEP and TIMSS scores that will be used to improve the accuracy of the predicted TIMSS state scores. Eight states were invited to participate in TIMSS separately from the nation at the 8th grade (Alabama, California, Colorado, Connecticut, Indiana, Massachusetts, Minnesota, and North Carolina) and one additional state, Florida, chose to participate itself. The actual TIMSS data from these nine states will be used to

[4]The primary purpose of stratification is to improve the precision of the survey estimates. If explicit stratification of the population is used, the units of interest (schools, for example) are sorted into mutually exclusive subgroups—strata. Units in the same stratum are as homogeneous as possible, and units in different strata are as heterogeneous as possible, with respect to the characteristics of interest to the survey. Separate samples are then selected from each stratum. In the case of implicit stratification, the units of interest are simply sorted with respect to one or more variables known to have a high correlation with the variable of interest. In this way, implicit stratification guarantees that the sample of units selected will be spread across the categories of the stratification variables.

[5]The Northeast region consists of Connecticut, Delaware, the District of Columbia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. The Central region consists of Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, Wisconsin, and South Dakota. The West region consists of Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oklahoma, Oregon, Texas, Utah, Washington, and Wyoming. The Southeast region consists of Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia.

[6]Four community types are distinguished: city of 250,000 or larger; suburb less than 250,000; town of 25,000 or more; rural metropolitan statistical area (MSA).

[7]In order to maximize response rates from both districts and schools it was necessary to begin the recruitment of both prior to the end of the 2009-10 school year. Since the 2011 NAEP sampling frame was not available until March 2010, it was necessary to base the TIMSS samples on the 2010 NAEP sampling frame.

[8]Since classrooms are sampled with equal probability within schools, small classrooms would have the same probability of selection as large classrooms. Selecting classrooms under these conditions would likely mean that student sample size would be reduced, and some instability in the sampling weights created. To avoid these problems, pseudo-classrooms are created for the purposes of classroom sampling, in which small classrooms are joined to reach a larger student count. These pseudo-classrooms are treated as single classes in the class sampling process. Following class sampling, the pseudo-classroom combinations are dissolved and the small classes involved retain their own identity. In this way, data on students, teachers, and classroom practices are linked in small classes in the same way as with larger classes.

[9]The classrooms selected could be pseudo-classrooms.

validate the predicted TIMSS average scores based on the linking study. The states invited to participate were selected based on state enrollment size and willingness to participate, as well as on their general NAEP performance (above or below the national average on NAEP), their previous experience in benchmarking to TIMSS, and their regional distribution.

To facilitate the linking study, NAEP added a science assessment at the national and state levels during the January–March 2011 collection period, with all states and the District of Columbia agreeing to participate. In addition, during the NAEP assessment window, a separate national sample of students was administered a set of "braided" booklets that included both NAEP and TIMSS blocks of items. During the NAEP assessment window, the braided booklets were designed to appear as similar as possible to a regular NAEP assessment booklet and were administered under the same conditions as NAEP.

During the TIMSS data collection period in spring 2011, TIMSS added state-level assessments in the nine states noted above as well as a separate national sample of students who were also administered a set of braided booklets that included both NAEP and TIMSS blocks of items. During the TIMSS assessment window, the braided booklets were designed to appear as similar as possible to a regular TIMSS assessment booklet and were administered under the same conditions as TIMSS. The inclusion of two additional samples of students who were administered the braided booklets during the NAEP and TIMSS assessment windows is similar to the braided booklet design used for recent NAEP studies to maintain NAEP trends in 2009 reading and 12th-grade mathematics.

Within each school, selected intact mathematics classes at 8th grade were randomly designated as a TIMSS class or a NAEP-TIMSS Linking Study class. The students in the linking study classes were administered the braided TIMSS-NAEP booklets. Those in the TIMSS classes were administered the regular TIMSS booklets.

The national and state samples were selected to coordinate the main NAEP and TIMSS samples. The data were collected by the regular NAEP field staff in winter and the regular TIMSS field staff in spring. The responses to the assessment questions were scored using the same scoring staff with the same training and quality control procedures that NAEP and TIMSS normally use. The results of the effort to link NAEP and TIMSS are to be released in a separate report from NCES.

Additional state samples

In addition to the states that participated at 8th grade, North Carolina also participated in TIMSS at 4th grade. North Carolina, like Florida at 8th grade, elected to participate in TIMSS on its own. The process for selecting the North Carolina school sample was the same as that employed in selecting the other state school samples (described below).

Selecting school samples for the states

The TIMSS state samples included only public schools. The school frame was identical to the national frame of public schools in those states. The state samples included the public schools in each state that were previously selected as part of the TIMSS national sample plus a supplement of schools. The sample target was 100 assessed classrooms. The target reference is classrooms because in the national design one class in each school is selected for TIMSS and one for the NAEP-TIMSS Linking Study. Thus, only one class from each of the national public schools was included in the state assessment. The supplemental sample of schools selected for each state followed the normal TIMSS procedure of selecting two classes per school. The additional number of schools needed in each state is then ([100 - # national public schools] / 2) plus an additional five schools per state to account for ineligible schools (schools with no students in the target grade).

The state sample was selected using a version of the Keyfitz procedure. Chowdhury, Chu, and Kaufman (2001) have described the implementation of the procedure. The method is generally used to minimize overlap but it can also be used to maximize overlap by ordering the rows in descending order of the response load indicator. By following the process outlined in table 2 of the paper, the rows in the table can be thought of as a hierarchy of selection preference where the top row maximizes the probability and the bottom row minimizes it. This property allowed us to maximize the overlap with the TIMSS national sample (in fact, select all national public schools) and minimize the overlap with the NAEP state operational public school sample or "Alpha sample."[10] This minimization was undertaken to reduce the burden for schools selected in the NAEP sample and to improve response rates. This was accomplished by partitioning the frame into the following three groups shown in order as in table 2 of the paper. The three groups are:

1. public schools selected for the TIMSS national sample (including schools also selected for the NAEP Alpha sample);
2. public schools not selected for either the TIMSS national or NAEP Alpha samples; and
3. public schools selected for the NAEP Alpha sample and not the TIMSS national sample.

The method guarantees all schools in group 1 will be selected with certainty since the probability of being selected for the validation sample is always larger than that for being selected for the national sample, because more schools were selected

[10]For a complete definition of the NAEP Alpha sample, see the NCES NAEP glossary at http://nces.ed.gov/nationsreportcard/glossary.asp#alpha_sample.

in each state sample (the national public schools plus a state supplement of public schools) than in the national sample with the frames being identical. The method minimized the overlap with schools in group 3 (NAEP Alpha sample) and selected the majority of the state supplement from schools in group 2.

U.S. TIMSS 4th-grade sample

School sample. The 4th-grade national school sample consisted of 450 schools (both public and private). As described previously, the joint administration of TIMSS and PIRLS at 4th grade required a larger sample of schools to ensure an adequate number of participating classes and students in both studies. Twelve ineligible schools and one excluded school were identified on the basis that they served special student populations, or had closed or altered their grade makeup since the sampling frame was developed. This left 437 schools eligible to participate, and 347 agreed to do so. The unweighted school response rate before substitution then was 79 percent. The analogous weighted school response rate was also 79 percent (see table A-1) and is given by the following formula:

$$\text{weighted school response rate before replacement} = \frac{\sum_{i \in Y} W_i E_i}{\sum_{i \in (Y \cup N)} W_i E_i}$$

where *Y* denotes the set of responding original-sample schools; *N* denotes the set of eligible non-responding original sample schools; Wi denotes the base weight for school *i;* *Wi* = 1/*Pi,* where *Pi* denotes the school selection probability for school *i;* and *Ei* denotes the enrollment size of age-eligible students, as indicated in the sampling frame.

In addition to the 347 participating schools from the original sample, 22 substitute schools participated for a total of 369 participating schools at the 4th grade in the United States (see table A-2). This gives a weighted (and unweighted) school participation rate after substitution of 84 percent (see table A-1).[11]

Classroom sample. Schools agreeing to participate in TIMSS were asked to list their 4th-grade mathematics classes as the basis for sampling at the classroom level. At this time, schools were given the opportunity to identify special classes–classes in which all or most of the students had intellectual or functional disabilities or were non-English language speakers. While these classes were regarded as eligible, the students as a group were treated as "excluded" since, in the opinion of the school, their disabilities or language capabilities would render meaningless their performance on the assessment. Fifty 4th-grade schools excluded classes and 681 students were excluded from participation in TIMSS as a result.

Prior to sampling, classes with fewer than 15 students were collapsed with other classes into what are called pseudo-classrooms. Creating pseudo-classrooms in this way ensured that all eligible classrooms in a school had at least 20 students. Up to four eligible classrooms were selected with classes being randomly assigned to TIMSS or PIRLS. In schools with only one classroom, this classroom was selected with certainty and randomly assigned to TIMSS or PIRLS. Some 1,257 classrooms were selected as a result of this process. All selected classrooms participated in TIMSS, yielding a classroom response rate of 100 percent (Martin et al. 2012, exhibit C.8).

Student sample. Schools were asked to list the students in each of the classrooms, along with the teachers who taught mathematics and science to these students. A total of 14,205 students were listed as a result. These students are identified by IEA as "sampled students in participating schools" (see table A-2).

This pool of students is reduced by within-school exclusions and withdrawals. At the time schools listed the students in the sampled classrooms, they had the opportunity to identify particular students who were not suited to take the test because of physical or intellectual disabilities (i.e., students with disabilities who had been mainstreamed) or because they were non-English-language speakers. Schools identified a total of 839 students they wished to have excluded from the assessment. By the time of the assessment a further 185 of the listed students had withdrawn from the school or classroom. In total, then, the pool of 14,205 sampled students was reduced by 1,024 students (839 excluded and 185 withdrawn) to yield 13,181 "eligible" students. The number of eligible students is used as the base for calculating student response rates (Martin et al. 2012, exhibit C.6).

The number of eligible students was further reduced on assessment day by 612 student absences, leaving 12,569 "assessed students" identified as having completed a TIMSS 2011 assessment booklet (see table A-2). IEA defines the student response rate as the number of students assessed as a percentage of the number of eligible students which, in this case, yields a weighted (and unweighted) student response rate of 95 percent (see table A-1).

Note that the 681 students excluded because whole classes were excluded do not figure in the calculation of student response rates. They do, however, figure in the calculation of the coverage of the International Target Population. Together, these 681 students excluded prior to classroom sampling, plus the 839 within-class exclusions, resulted in an overall student exclusion rate of 7.0 percent (see table A-1 and Martin et al. 2012, exhibit C.3). The reported coverage

[11] Substitute schools are matched pairs and do not have an independent probability of selection. NCES standards (Standard 1-3-8) indicate that, in these circumstances, response rates should be calculated without including substitute schools (National Center for Education Statistics 2002). TIMSS response rates denoted as "before replacement" conform to this standard. TIMSS response rates denoted as "after replacement" are not consistent with NCES standards since, in the calculation of these rates, substitute schools are treated as the equivalent of sampled schools.

of the International Target Population then is 93 percent (see Martin et al. 2012, exhibit C.3).

Combined participation rates. For the results for an education system to be included in the TIMSS international report without a response rate annotation, the IEA requires a "combined" or overall response rate—expressed as the product of (a) the (unrounded) weighted school response rate without substitute schools and (b) the (unrounded) weighted student response rate—of at least 75 percent (after rounding to the nearest whole percent). The overall response rate for the United States, 75.72 percent without substitute schools, meets this requirement. However, the United States did include substitute schools because its school-level response rate was less than 85 percent, and, absent advance knowledge of the student-level response rate, introducing substitute schools was a prudent approach to take. For the results of an education system to be included in the TIMSS international report without a student inclusion annotation, the IEA requires a student inclusion rate of at least 95 percent. Because 7 percent of the 4th-grade student population was excluded in the United States, the overall U.S. student inclusion rate was 93 percent. For this reason, the U.S. 4th-grade results in the TIMSS international report carry a coverage annotation indicating that coverage of the defined student population was less than the IEA standard of 95 percent.

Tables A-1 and A-2 are extracts from the international report exhibits noted above and are designed to summarize information on school and student responses rates and coverage of the 4th- and 8th-grade target populations in each nation.

U.S. TIMSS 8th-grade sample

School sample. The 8th-grade national school sample consisted of 600 schools. Twenty-two ineligible original schools and four excluded schools were identified on the basis that they served special student populations, or had closed or altered their grade makeup since the sampling frame was developed. This left 574 schools eligible to participate and 499 agreed to do so. The unweighted original school response rate before substitution then was 87 percent. The analogous weighted school response rate was 87 percent (see table A-1).

In addition to the 499 participating schools from the original sample, 2 substitute schools participated for a total of 501 participating schools at the 8th grade in the United States (see table A-2). This gives a weighted (and unweighted) school participation rate after substitution of 87 percent (see table A-1).[12]

Classroom sample. Schools agreeing to participate were asked to list their 8th-grade mathematics classes as the basis for sampling at the classroom level. At this time, schools were given the opportunity to identify special classes—classes in which all or most of the students had intellectual or functional disabilities or were non-English-language speakers. While these classes were regarded as eligible, the students as a group were treated as "excluded" since, in the opinion of the school, their disabilities or language capabilities would render meaningless their performance on the assessment. A total of 223 schools excluded classrooms from participation in TIMSS. This resulted in the exclusion of 4,650 students.

Classrooms with fewer than 15 students were collapsed into pseudo-classrooms prior to sampling so that each eligible classroom in a school had at least 20 students. Two eligible classrooms were selected per school where possible. In schools where there was only one eligible classroom, this classroom was selected with certainty and was randomly assigned to TIMSS or the TIMSS-NAEP Linking Study. All selected classrooms participated in TIMSS yielding a classroom response rate of 100 percent (Martin et al. 2012, exhibit C.9).

Student sample. Schools were asked to list the students in each of the sampled mathematics classrooms, along with the teachers who taught mathematics and science to these students. A total of 11,864 students were listed as being in the selected classrooms. These students are identified by IEA as "sampled students in participating schools" (see table A-2).

This pool of students is reduced by within-school exclusions and withdrawals. At the time schools listed the students in sampled classrooms, they had the opportunity to identify particular students who were not suited to take the test because of physical or intellectual disabilities (i.e., students with disabilities who had been mainstreamed) or because they were non-English language speakers. Schools identified a total of 398 students they wished to have excluded from the assessment; and, by the time of the assessment, a further 302 of the listed students had withdrawn from the school or classroom. In total then, the pool of 11,864 sampled students was reduced by 700 students (398 excluded and 302 withdrawn) to yield 11,164 "eligible" students. The number of eligible students is used as the base for calculating student response rates (Martin et al. 2012, exhibit C.7).

[12]Substitute schools are matched pairs and do not have an independent probability of selection. NCES standards (Standard 1-3-8) indicate that, in these circumstances, response rates should be calculated without including substitute schools (National Center for Education Statistics 2002). TIMSS response rates denoted as "before replacement" conform to this standard. TIMSS response rates denoted as "after replacement" are not consistent with NCES standards since, in the calculation of these rates, substitute schools are treated as the equivalent of sampled schools.

Table A-1. Coverage of target populations, school participation rates, and student response rates, by grade and education system: 2011

	Grade 4						
Education system	Average age at time of tesing	Percentage of international desired population coverage	National desired population overall exclusion rate	Weighted school participation rate before substitution	Weighted school participation rate after substitution	Weighted student response rate	Combined weighted school participation and student response rate with substitute schools
Armenia	10	100	2	100	100	98	98
Australia	10	100	4	96	98	95	93
Austria	10	100	5	100	100	98	98
Azerbaijan	10	100	7	84	100	100	100
Bahrain	10	100	1	92	92	98	90
Belgium (Flemish)-BEL	10	100	5	76	95	98	92
Chile	10	100	4	86	99	96	95
Chinese Taipei-CHN	10	100	1	100	100	99	99
Croatia	11	100	8	99	100	95	95
Czech Republic	10	100	5	90	99	95	94
Denmark	11	100	6	79	92	95	87
England-GBR	10	100	2	81	83	94	78
Finland	11	100	3	97	99	96	96
Georgia	10	92	5	97	98	99	96
Germany	10	100	2	96	99	96	95
Hong Kong-CHN	10	100	9	87	88	93	82
Hungary	11	100	4	98	99	97	96
Iran, Islamic Rep. of	10	100	5	100	100	99	99
Ireland	10	100	3	97	99	95	95
Italy	10	100	4	81	98	97	95
Japan	11	100	3	96	99	97	97
Kazakhstan	10	100	6	99	100	99	99
Korea, Rep. of	10	100	3	100	100	98	98
Kuwait	10	78	0	99	99	94	91
Lithuania	11	93	6	94	100	94	94
Malta	10	100	4	100	100	95	95
Morocco	11	100	2	100	100	97	96
Netherlands	10	100	4	49	82	97	79
New Zealand	10	100	5	83	96	94	90
Northern Ireland-GBR	10	100	3	62	85	93	79
Norway	10	100	4	57	82	85	70
Oman	10	100	1	98	98	98	96
Poland	10	100	4	100	100	96	96
Portugal	10	100	3	87	98	94	92
Qatar	10	100	6	100	100	99	99
Romania	11	100	4	99	100	98	97
Russian Federation	11	100	5	100	100	98	98
Saudi Arabia	10	100	2	95	100	99	99
Serbia	11	100	9	97	100	97	97
Singapore	10	100	6	100	100	96	96
Slovak Republic	10	100	5	95	99	96	96
Slovenia	10	100	3	96	97	97	94
Spain	10	100	5	96	99	97	97
Sweden	11	100	4	97	99	92	91
Thailand	11	100	2	85	100	99	99
Tunisia	10	100	2	100	100	99	99
Turkey	10	100	3	97	100	98	98
United Arab Emirates	10	100	3	100	100	97	97
United States	10	100	7	79	84	95	80
Yemen	11	100	4	99	99	97	95

See notes at end of table.

Table A-1. Coverage of target populations, school participation rates, and student response rates, by grade and education system: 2011—Continued

	Grade 4						
Benchmarking education systems	Average age at time of tesing	Percentage of international desired population coverage	National desired population overall exclusion rate	Weighted school participation rate before substitution	Weighted school participation rate after substitution	Weighted student response rate	Combined weighted school participation and student response rate with substitute schools
Alberta-CAN	10	100	8	98	99	96	95
Ontario-CAN	10	100	5	97	98	96	94
Quebec-CAN	10	100	4	95	96	95	91
Abu Dhabi-UAE	10	100	3	99	99	98	97
Dubai-UAE	10	100	5	100	100	96	96
Florida-USA	10	89	12	96	96	95	91
North Carolina-USA	10	93	10	94	94	95	89

See notes at end of table.

Table A-1. Coverage of target populations, school participation rates, and student response rates, by grade and education system: 2011—Continued

	Grade 8						
Education system	Average age at time of tesing	Percentage of international desired population coverage	National desired population overall exclusion rate	Weighted school participation rate before substitution	Weighted school participation rate after substitution	Weighted student response rate	Combined weighted school participation and student response rate with substitute schools
Armenia	15	100	2	100	100	97	97
Australia	14	100	3	96	98	90	88
Bahrain	14	100	2	99	99	98	97
Chile	14	100	3	88	99	95	95
Chinese Taipei-CHN	14	100	1	100	100	99	99
England-GBR	14	100	2	75	79	89	70
Finland	15	100	3	97	98	95	93
Georgia	14	93	5	97	98	98	97
Ghana	16	100	1	100	100	97	97
Hong Kong-CHN	14	100	5	77	78	96	75
Hungary	15	100	4	98	99	96	95
Indonesia	14	100	3	100	100	96	96
Iran, Islamic Rep. of	14	100	2	100	100	99	99
Israel	14	100	23	94	100	92	92
Italy	14	100	5	83	97	96	93
Japan	15	100	3	85	92	94	87
Jordan	14	100	0	100	100	96	96
Kazakhstan	15	100	5	99	100	98	98
Korea, Rep. of	14	100	2	100	100	99	99
Lebanon	14	100	1	90	98	96	94
Lithuania	15	93	5	92	99	93	92
Macedonia, Rep. of	15	100	3	100	100	95	95
Malaysia	14	100	0	100	100	98	98
Morocco	15	100	0	100	100	94	94
New Zealand	14	100	3	87	98	90	88
Norway	14	100	2	89	89	94	84
Oman	14	100	1	99	99	98	97
Palestinian Nat'l Authority	14	100	2	100	100	98	98
Qatar	14	100	5	99	99	99	99
Romania	15	100	1	99	100	99	99
Russian Federation	15	100	6	100	100	98	98
Saudi Arabia	14	100	1	98	100	98	98
Singapore	14	100	6	100	100	95	95
Slovenia	14	100	2	96	98	94	92
Sweden	15	100	5	97	98	94	92
Syrian Arab Republic	14	100	2	99	99	93	92
Thailand	14	100	2	92	100	99	99
Tunisia	14	100	0	99	99	97	97
Turkey	14	100	2	99	100	97	97
Ukraine	14	100	3	98	100	98	98
United Arab Emirates	14	100	3	100	100	97	97
United States	14	100	7	87	87	94	81

See notes at end of table.

Table A-1. Coverage of target populations, school participation rates, and student response rates, by grade and education system: 2011—Continued

	Grade 8						
Benchmarking education systems	Average age at time of tesing	Percentage of international desired population coverage	National desired population overall exclusion rate	Weighted school participation rate before substitution	Weighted school participation rate after substitution	Weighted student response rate	Combined weighted school participation and student response rate with substitute schools
Alberta-CAN	14	100	7	91	99	93	92
Ontario-CAN	14	100	6	97	98	95	93
Quebec-CAN	14	100	5	96	96	93	88
Abu Dhabi-UAE	14	100	2	99	99	97	96
Dubai-UAE	14	100	4	99	99	96	95
Alabama-USA	14	92	5	92	92	92	84
California-USA	14	91	6	85	88	94	82
Colorado-USA	14	94	4	84	89	94	84
Connecticut-USA	14	90	9	100	100	94	94
Florida-USA	14	89	7	94	94	91	84
Indiana-USA	14	90	6	94	97	96	93
Massachusetts-USA	14	89	8	100	100	96	96
Minnesota-USA	14	90	4	91	98	95	94
North Carolina-USA	14	93	11	98	98	95	93

NOTE: Educations systems in the Southern hemisphere administered “TIMSS 2011” in the fall of 2010 while those in the Northern hemisphere administered the assessment in the spring of 2011. Italics indicate participants identified and counted in this report as an education system and not as a separate country. The international desired population refers to the sample and not the responding schools, classes, and students.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Table A-2. Total number of schools and students, by grade and education system: 2011

	Grade 4						
Education system	Schools in original sample	Eligible schools in original sample	Schools in original sample that participated	Substitute schools	Total schools that participated	Sampled students in participating schools	Students assessed
Armenia	150	150	150	0	150	5,292	5,146
Australia	290	284	275	5	280	6,709	6,146
Austria	160	158	158	0	158	4,976	4,668
Azerbaijan	170	169	142	27	169	5,098	4,882
Bahrain	174	172	159	0	159	4,213	4,083
Belgium (Flemish)-BEL	156	150	114	28	142	5,219	4,849
Chile	203	202	169	31	200	6,010	5,585
Chinese Taipei-CHN	150	150	150	0	150	4,376	4,284
Croatia	152	152	150	2	152	5,097	4,584
Czech Republic	180	178	161	16	177	4,895	4,578
Denmark	240	235	186	30	216	4,452	3,987
England-GBR	150	150	122	3	125	3,689	3,397
Finland	150	146	141	4	145	4,917	4,638
Georgia	180	177	172	1	173	4,958	4,799
Germany	200	199	190	7	197	4,229	3,995
Hong Kong-CHN	154	154	134	2	136	4,330	3,957
Hungary	150	150	146	3	149	5,488	5,204
Iran, Islamic Rep. of	250	244	244	0	244	5,932	5,760
Ireland	152	151	147	3	150	4,836	4,560
Italy	205	205	166	36	202	4,529	4,200
Japan	150	150	144	5	149	4,595	4,411
Kazakhstan	150	149	147	2	149	4,521	4,382
Korea, Rep. of	150	150	150	0	150	4,494	4,334
Kuwait	150	150	148	0	148	4,431	4,142
Lithuania	160	154	145	9	154	5,140	4,688
Malta	99	96	96	0	96	3,958	3,607
Morocco	289	287	286	0	286	8,414	7,841
Netherlands	151	148	75	53	128	3,461	3,229
New Zealand	189	189	154	26	180	6,172	5,572
Northern Ireland-GBR	160	160	100	36	136	3,942	3,571
Norway	150	145	84	35	119	3,881	3,121
Oman	338	333	327	0	327	10,840	10,411
Poland	150	150	150	0	150	5,316	5,027
Portugal	150	150	132	15	147	4,384	4,042
Qatar	175	167	166	0	166	4,394	4,117
Romania	150	148	147	1	148	4,879	4,673
Russian Federation	202	202	202	0	202	4,693	4,467
Saudi Arabia	175	171	163	8	171	4,625	4,515
Serbia	160	156	152	4	156	4,603	4,379
Singapore	176	176	176	0	176	6,687	6,368
Slovak Republic	200	198	187	10	197	5,933	5,616
Slovenia	202	201	193	2	195	4,674	4,492
Spain	152	152	147	4	151	4,461	4,183
Sweden	161	153	148	4	152	5,235	4,663
Thailand	168	168	143	25	168	4,556	4,448
Tunisia	222	222	222	0	222	5,057	4,912
Turkey	260	257	251	6	257	7,905	7,479
United Arab Emirates	478	460	459	0	459	15,428	14,720
United States	450	437	347	22	369	14,205	12,569
Yemen	223	218	216	0	216	8,794	8,058

See notes at end of table.

Table A-2. Total number of schools and students, by grade and education system: 2011—Continued

Benchmarking education systems	Grade 4						
	Schools in original sample	Eligible schools in original sample	Schools in original sample that participated	Substitute schools	Total schools that participated	Sampled students in participating schools	Students assessed
Alberta-CAN	150	144	141	2	143	4,086	3,645
Ontario-CAN	150	149	145	1	146	5,022	4,570
Quebec-CAN	200	197	189	1	190	4,529	4,235
Abu Dhabi-UAE	168	165	164	0	164	4,308	4,164
Dubai-UAE	152	139	139	0	139	6,553	6,151
Florida-USA	81	80	77	0	77	3,121	2,661
North Carolina-USA	49	49	46	0	46	2,104	1,792

See notes at end of table.

Table A-2. Total number of schools and students, by grade and education system: 2011—Continued

	Grade 8						
Education system	Schools in original sample	Eligible schools in original sample	Schools in original sample that participated	Substitute schools	Total schools that participated	Sampled students in participating schools	Students assessed
Armenia	153	153	153	0	153	6,057	5,846
Australia	290	287	276	1	277	9,007	7,556
Bahrain	97	96	95	0	95	4,960	4,640
Chile	197	196	166	27	193	6,290	5,835
Chinese Taipei-CHN	150	150	150	0	150	5,166	5,042
England-GBR	150	150	113	5	118	4,382	3,842
Finland	150	148	143	2	145	4,549	4,266
Georgia	180	175	171	1	172	4,779	4,563
Ghana	163	161	161	0	161	8,073	7,323
Hong Kong-CHN	150	150	116	1	117	4,261	4,015
Hungary	150	147	144	2	146	5,489	5,178
Indonesia	154	153	153	0	153	6,201	5,795
Iran, Islamic Rep. of	250	238	237	1	238	6,264	6,029
Israel	152	151	143	8	151	5,174	4,699
Italy	204	204	166	31	197	4,379	3,979
Japan	150	150	128	10	138	4,747	4,414
Jordan	232	230	230	0	230	8,439	7,694
Kazakhstan	150	147	146	1	147	4,551	4,390
Korea, Rep. of	150	150	150	0	150	5,315	5,166
Lebanon	150	150	136	11	147	4,231	3,974
Lithuania	150	142	132	9	141	5,285	4,747
Macedonia, Rep. of	150	150	150	0	150	4,360	4,062
Malaysia	180	180	180	0	180	6,209	5,733
Morocco	285	280	279	0	279	9,869	8,986
New Zealand	162	162	141	17	158	6,079	5,336
Norway	150	150	134	0	134	4,229	3,862
Oman	338	333	323	0	323	9,947	9,542
Palestinian Nat'l Authority	203	201	201	0	201	8,069	7,812
Qatar	113	110	109	0	109	4,641	4,422
Romania	150	147	145	2	147	5,704	5,523
Russian Federation	210	210	210	0	210	5,146	4,893
Saudi Arabia	154	153	150	3	153	4,477	4,344
Singapore	165	165	165	0	165	6,314	5,927
Slovenia	191	191	183	3	186	4,722	4,415
Sweden	159	156	152	1	153	6,210	5,573
Syrian Arab Republic	150	150	148	0	148	4,756	4,413
Thailand	172	172	160	12	172	6,404	6,124
Tunisia	217	211	207	0	207	5,464	5,128
Turkey	240	239	237	2	239	7,348	6,928
Ukraine	150	148	146	2	148	3,491	3,378
United Arab Emirates	477	460	458	0	458	14,716	14,089
United States	600	574	499	2	501	11,864	10,477

See notes at end of table.

Table A-2. Total number of schools and students, by grade and education system: 2011—Continued

	Grade 8						
Benchmarking education systems	Schools in original sample	Eligible schools in original sample	Schools in original sample that participated	Substitute schools	Total schools that participated	Sampled students in participating schools	Students assessed
Alberta-CAN	150	147	133	12	145	5,579	4,799
Ontario-CAN	150	146	142	1	143	5,198	4,756
Quebec-CAN	200	198	189	0	189	6,879	6,149
Abu Dhabi-UAE	170	167	166	0	166	4,513	4,373
Dubai-UAE	143	131	130	0	130	5,915	5,571
Alabama-USA	63	60	55	0	55	2,414	2,113
California-USA	94	93	79	3	82	2,898	2,614
Colorado-USA	60	60	50	3	53	2,395	2,167
Connecticut-USA	63	62	62	0	62	2,356	2,099
Florida-USA	65	64	60	0	60	1,986	1,712
Indiana-USA	62	58	55	1	56	2,501	2,260
Massachusetts-USA	58	56	56	0	56	2,296	2,075
Minnesota-USA	60	56	51	4	55	2,720	2,500
North Carolina-USA	62	60	59	0	59	2,434	2,103

NOTE: Educations systems in the Southern hemisphere administered “TIMSS 2011” in the fall of 2010 while those in the Northern hemisphere administered the assessment in the spring of 2011. Italics indicate participants identified and counted in this report as an education system and not as a separate country
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

The number of eligible students was further reduced on assessment day by 687 student absences, leaving 10,477 "assessed students" identified as having completed a TIMSS 2011 assessment booklet (see table A-2). The IEA defines the student response rate as the number of students assessed as a percentage of the number of eligible students which, in this case yields a weighted (and unweighted) student response rate of 94 percent (see table A-1).

Note that the 4,650 students excluded because whole classes were excluded do not figure in the calculation of student response rates. They do, however, figure in the calculation of the coverage of the International Target Population. Together, these 4,650 students excluded prior to classroom sampling, plus the 398 within-class exclusions resulted in an overall student exclusion rate of 7.2 percent (see table A-1 and Martin et al. 2012, exhibit C.3). The reported coverage of the International Target Population then is 100 percent (see Martin et al. 2012, exhibit C.3).

Combined participation rates. TIMSS' combined school, classroom, and student weighted response rate standard of 75 percent was met: the weighted and unweighted product of the separate U.S. response rates (81 percent) exceeded this 75 percent standard (see table A-1). However, the United States did include substitute schools because its school-level response rate was less than 85 percent, and, absent advance knowledge of the student-level response rate, introducing substitute schools was a prudent approach to take. Because 7 percent of the 8th-grade student population was excluded in the United States, the overall U.S. student inclusion rate was 93 percent. For this reason, the U.S. 8th-grade results in the TIMSS international report carry a coverage annotation indicating that coverage of the defined student population was less than the IEA standard of 95 percent. Table A-2 summarizes information on the coverage of the 8th-grade target populations in each participating education system.

Nonresponse bias in the U.S. TIMSS samples

NCES standards require a nonresponse bias analysis if the school-level response rate falls below 85 percent of the sampled schools (standard 2-2-2; NCES Education Statistics 2002), as it did for the 4th-grade sample. As a consequence, a nonresponse bias analysis was initiated and took a form similar to that adopted for TIMSS 2003 (Ferraro and Van de Kerckhove 2006). A full report of this study will be included in a technical report to be released with the U.S. national TIMSS dataset. The state samples were sufficient enough that none required a nonresponse bias analysis.

Three methods were chosen to perform this analysis. The first method focused exclusively on the sampled schools and ignored substitute schools. The schools were weighted by their school base weights, excluding any nonresponse adjustment factor. The second method focused on *sampled schools plus substitute schools*, treating as nonrespondents those schools from which a final response was not received. Again, schools were weighted by their base weights, with the base weight for each substitute school set to the base weight of the original school that it replaced. The third method repeated the analyses from the second method using nonresponse adjusted weights.[13]

In order to compare TIMSS respondents and nonrespondents, it was necessary to match the sample of schools back to the sample frame to identify as many characteristics as possible that might provide information about the presence of nonresponse bias.[14] The characteristics available for analysis in the sampling frame were taken from the CCD for public schools, and from the PSS for private schools. For categorical variables, the distribution of the characteristics for respondents was compared with the distribution for all schools. The hypothesis of independence between a given school characteristic and the response status (whether or not the school participated) was tested using a Rao-Scott modified chi-square statistic. For continuous variables, summary means were calculated and the difference between means was tested using a *t* test. Note that this procedure took account of the fact that the two samples in question were not independent samples, but in fact the responding sample was a subsample of the full sample. This effect was accounted for in calculating the standard error of the difference. Note also that in those cases where both samples were weighted using just the base weights, the test is exactly equivalent to testing that the mean of the respondents was equal to the mean of the nonrespondents.

In addition, multivariate logistic regression models were set up to identify whether any of the school characteristics were significant in predicting response status when the effects of all potential influences available for modeling were considered simultaneously.

Public and private schools were modeled together using the following variables:[15] community type (city, suburban, town, and rural); control of school (public or private); Census region (Northeast, Southeast, Central, and West); poverty level (percentage of students in school eligible for free or reduced-

[13]A detailed treatment of the meaning and calculation of sampling weights, including the nonresponse adjustment factors, is provided in *TIMSS and PIRLS Methods and Procedures* (Martin and Mullis 2011).

[14]Comparing characteristics for respondents and nonrespondents is not always a good measure of nonresponse bias if the characteristics are either unrelated or weakly related to more substantive items in the survey. Nevertheless, this is often the only approach available.

[15]NAEP region and community type were dummy coded for the purposes of these analyses. In the case of NAEP region, "West" was used as the reference group. For community type, "urban fringe/large town" was chosen as the reference group.

price lunch);[16] number of students enrolled in 4th-grade; total number of students; and, percentage minority students.[17]

Results for the original sample of schools. In the analyses for the original sample of schools, all substituted schools were treated as nonresponding schools. The results of these analyses follow.

Fourth grade. In the investigation into nonresponse bias at the school level for TIMSS 4th-grade schools, comparisons between schools in the eligible sample and participating schools showed that there was no relationship between response status and the majority of school characteristics available for analysis. In separate variable-by-variable bivariate analyses, six variables were found to be statistically significantly related to participation in the bivariate analysis: school control, community type, 4th-grade enrollment, the percentage of Black, non-Hispanic students, the percentage of Hispanic students, and the percentage of American Indian or Alaska Native students. Although each of these findings indicates some potential for nonresponse bias, when all of these factors were considered simultaneously in a regression analysis, private schools, the South (as a region), total school enrollment, 4th-grade enrollment, and the percentage of American Indian or Alaska Native students and Hispanics were significant predictors of participation. The second model showed that private schools, total school enrollment, and 4th-grade enrollment were significant predictors of participation.

Results for the final sample of schools. In the analyses for the final sample of schools, all substitute schools were included with the original schools as responding schools, leaving nonresponding schools as those for which no assessment data were available. The results of these analyses follow and are somewhat more complicated than the analyses for the original sample of schools.

Fourth grade. The bivariate results for the final sample of 4th-grade schools indicated that five of the six variables remained statistically significant in the bivariate analysis: school control, 4th-grade enrollment, the percentage of Black, non-Hispanic students, the percentage of Hispanic students, and the percentage of American Indian or Alaska Native students. When all of these factors were considered simultaneously in a regression analysis, private schools, total school enrollment, and the percentage of American Indian or Alaska Native students remained significant predictors of participation.

For the final sample of schools in 4th grade, with school nonresponse adjustments applied to the weights, only two variables were statistically significant in the bivariate analysis: region and the percentage of Black, non-Hispanic students.[18] The multivariate regression analysis cannot be conducted after the school nonresponse adjustments are applied to the weights.

These results suggest that there is some potential for nonresponse bias in the U.S. 4th-grade original sample based on the characteristics studied. It also suggests that, while there is no evidence that the use of substitute schools reduced the potential for bias, it has not added to it substantially and the inclusion of the replacement schools increased sample available for analyses. The application of school nonresponse adjustments substantially reduced the potential for bias.

Test development

TIMSS is a cooperative effort involving representatives from every country participating in the study. For TIMSS 2011, the test development effort began with a review and revision of the frameworks that are used to guide the construction of the assessment (Mullis et al. 2009). The frameworks were updated to reflect changes in the curriculum and instruction of participating countries and education systems. Extensive input from experts in mathematics and science education, assessment, and curriculum, and representatives from national educational centers around the world contributed to the final shape of the frameworks. Maintaining the ability to measure change over time was an important factor in revising the frameworks.

As part of the TIMSS dissemination strategy, approximately one-half of the 2007 assessment items were released for public use. To replace assessment items that had been released, participants submitted items for review by subject-matter specialists, and additional items were written by the IEA Science and Mathematics Review Committee in consultation with item-writing specialists in various countries

[16]The measure of school poverty is based on the proportion of students in a school eligible for the free or reduced-price lunch (FRPL) program, a federally assisted meal program that provides nutritionally balanced, low-cost or free lunches to eligible children each school day. For the purposes of the nonresponse bias analyses, schools were classified as "low poverty" if less than 50 percent of the students were eligible for FRPL and as "high poverty" if 50 percent or more of the students were eligible. Since the nonresponse bias analyses involve both participating and nonparticipating schools, they are based, out of necessity, on data from the sampling frame. TIMSS data are not available for nonparticipating schools. The school frame data are derived from the CCD and PSS. The CCD data provide information on the percentage of students in each school who are eligible for free or reduced-price lunch, but are limited to public schools. The PSS data do not provide the same information for private schools. In the interest of retaining all of the schools and students in these analyses, private schools were assumed to be low-poverty schools–that is, they were assumed to be schools in which less than 50 percent of students were eligible for FRPL. Separate analyses of the TIMSS data for participating private schools suggest the reasonableness of this assumption. Of the 21 grade 4 private schools, only one reports having 50 percent or more of students eligible for FRPL.

[17]Two forms of this school attribute were used in the analyses. In the bivariate analyses the percentage of each racial/ethnic group was related separately to participation status. In the logistic regression analyses a single measure was used to characterize each school, namely, "percentage of minority students."

[18]The international weighting procedures created a nonresponse adjustment class for each explicit stratum; see *TIMSS and PIRLS Methods and Procedures* (Mullis et al. 2011) for details. In the case of the U.S. 4th-grade sample, there was no explicit stratification and thus a single adjustment class. The procedures could not be varied for individual countries to account for any specific needs. Therefore, the U.S. nonresponse bias analyses could have no influence on the weighting procedures and were undertaken after the weighting process was complete.

Table A-3. Number and percentage distribution of new and trend mathematics and science items in the TIMSS assessment, by grade and domain: 2011

	Grade 4					
	All items		New items		Trend items	
Content domain	Number	Percent	Number	Percent	Number	Percent
Mathematics	175	100	72	100	103	100
Number	88	50	36	50	52	50
Geometric shapes and measures	61	35	25	35	36	35
Data display	26	15	11	15	15	15
Science	172	100	72	100	100	100
Life science	75	44	33	46	42	42
Physical science	63	37	25	35	38	38
Earth science	34	20	14	19	20	20

	Grade 8					
	All items		New items		Trend items	
Content domain	Number	Percent	Number	Percent	Number	Percent
Mathematics	217	100	91	100	126	100
Number	61	28	31	34	20	16
Algebra	70	32	23	25	47	37
Geometry	43	20	18	20	25	20
Data and chance	43	20	19	21	24	19
Science	217	100	92	100	125	100
Biology	79	36	33	36	46	37
Chemistry	44	20	19	21	25	20
Physics	55	25	22	24	33	26
Earth science	39	18	18	20	21	17

NOTE: The percentages in this table represent the number of items and not the number of score points. Some constructed-response items are worth more than one score point. For the percentages of score points, see table 2. Details may not sum to 100 percent due to rounding.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

and education systems to ensure that the content, as explicated in the frameworks, was covered adequately. Items were reviewed by an international Science and Mathematics Item Review Committee and field-tested in most of the participating countries. Results from the field test were used to evaluate item difficulty, how well items discriminated between high- and low-performing students, the effectiveness of distracters in multiple-choice items, scoring suitability and reliability for constructed-response items, and evidence of bias toward or against individual countries or in favor of boys or girls. As a 4th-grade result of this review, 72 new 4th-grade mathematics and 72 new 4th-grade science items were selected for inclusion in the international assessment. In total, 175 mathematics and 172 science items were included in the 4th-grade TIMSS assessment booklets. At the 8th grade, the review of the item statistics from the field test led to the inclusion of 91 new 8th-grade mathematics and 92 new 8th-grade science items in the assessment. In total, 217 mathematics and 217 science items were included in the 8th-grade TIMSS assessment booklets. More detail on the distribution of new and trend items is included in table A-3.

Design of instruments

TIMSS 2011 included booklets containing assessment items as well as self-administered background questionnaires for principals, teachers, and students.

Assessment booklets

The assessment booklets were constructed such that not all of the students responded to all of the items. This is consistent with other large-scale assessments, such as NAEP. To keep the testing burden to a minimum, and to ensure broad subject-matter coverage, TIMSS used a rotated block design that included both mathematics and science items. That is, students encountered both mathematics and science items during the assessment.

In 2011, the 4th-grade assessment consisted of 14 booklets, each requiring approximately 72 minutes. The assessment was given in two 36-minute parts, with a 5- to 10-minute break in between. The student questionnaire was given after the second part of the assessment. Although it was untimed, it was allotted approximately 30 minutes for response time.

The 14 booklets were rotated among students, with each participating student completing 1 booklet only. The mathematics and science items were each assembled separately into 14 blocks, or clusters, of items. Each of the 14 TIMSS 2011 booklets contained 4 blocks in total. Each block contained either mathematics items or science items only and each block occurred twice across the 14 books. For each subject, the secure, or trend, items used in prior assessments were included in 8 blocks, with the other 6 blocks containing new items.

Table A-4. Number of mathematics and science items in the TIMSS grade 4 and grade 8 assessments, by type and content domain: 2011

	Grade 4			Grade 8		
		Response type			Response type	
Content domain	Total	Multiple choice	Constructed response	Total	Multiple choice	Constructed response
Total	**347**	**186**	**161**	**434**	**228**	**206**
Mathematics	175	93	82	217	118	99
Number	88	42	46	61	31	30
Geometric shapes and measures	61	38	23			
Algebra				70	37	33
Geometry				43	25	18
Data display	26	13	13			
Data and chance				43	25	18
Science	172	93	79			
Life science	75	36	39			
Physical science	63	37	26			
Earth science	34	20	14			
Science				217	110	107
Biology				79	38	41
Chemistry				44	22	22
Physics				55	29	26
Earth science				39	21	18

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

The 2011 8th-grade assessment followed the same pattern and consisted of 14 booklets, each requiring approximately 90 minutes of response time. The assessment was given in two 45-minute parts, with a 5- to 10-minute break in between. As in 4th grade, the student questionnaire was given after the second part of the assessment, and was allotted approximately 30 minutes of response time. The 14 booklets were rotated among students, with each participating student completing 1booklet only. The mathematics and science items were assembled into 14 blocks, or clusters, of items. Each block contained either mathematics items or science items only. For each subject, the secure, or trend, items used in prior assessments were included in 8 blocks, with the other 6 blocks containing new items. Each of the 14 TIMSS 2011 booklets contained 4 blocks in total. The TIMSS booklets administered in the state samples were exactly the same as those administered in the national sample.

As part of the design process, it was necessary to ensure that the booklets showed a distribution across the mathematics and science content domains as specified in the frameworks. The number of mathematics and science items in the 4th- and 8th-grade TIMSS 2011 assessments is shown in table A-4.

Background questionnaires

As in prior administrations, TIMSS 2011 included self-administered questionnaires for principals, teachers, and students. To create the questionnaires for 2011, the 2007 versions were reviewed extensively by the national research coordinators from the participating countries and education systems as well as a Questionnaire Item Review Committee (QIRC). The QIRC comprises 10–12 experienced National Research Coordinators (NRCs) from different participating countries and education systems who have analyzed TIMSS data and use it in their own countries or education system. The QIRC review resulted in items being deleted or revised, and the addition of several new ones. Like the assessment items, all questionnaire items were field-tested and the results reviewed carefully. As a result, some of the questionnaire items needed to be revised prior to their inclusion in the final questionnaires. The questionnaires requested information to help provide a context for the performance scores, focusing on such topics as students' attitudes and beliefs about learning, their habits and homework, and their lives both in and outside of school; teachers' attitudes and beliefs about teaching and learning, teaching assignments, class size and organization, instructional practices, and participation in professional development activities; and principals' viewpoints on policy and budget responsibilities, curriculum and instruction issues and student behavior, as well as descriptions of the organization of schools and courses. For 2011, online versions of the school and teacher questionnaires were offered to respondents as the primary mode of data collection. Detailed results from the student, teacher, and school surveys are not discussed in this report but are available in the two international reports: the *TIMSS 2011 International Mathematics Report* (Mullis et al. 2012) and *TIMSS 2011 International Science Report* (Martin et al. 2012).

Translation

Source versions of all instruments (assessment booklets, questionnaires, and manuals) were prepared in English and translated into the primary language or languages of instruction in each education system. In addition, it was sometimes necessary to adapt the instrument for cultural purposes, even in countries and education systems that use English as the primary language of instruction. All adaptations were reviewed and approved by the International Study Center to ensure they did not change the substance or intent of the question or answer choices. For example, proper names were sometimes changed to names that would be more familiar to students (e.g., Marja-leena to Maria).

Each participant prepared translations of the instruments according to translation guidelines established by the International Study Center. Adaptations to the instruments were documented by each participant and submitted for review. The goal of the translation guidelines was to produce translated instruments of the highest quality that would provide comparable data across participants.

Translated instruments were verified by an independent, professional translation agency prior to final approval and printing of the instruments. Participating education systems were required to submit copies of the final printed instruments to the International Study Center. Further details on the translation process can be found in the *TIMSS and PIRLS Methods and Procedures* (Martin and Mullis 2011).

Recruitment, test administration, and quality assurance

TIMSS 2011 emphasized the use of standardized procedures for all participants. Each participating country and education system collected its own data, based on comprehensive manuals and training materials provided by the international project team to explain the survey's implementation, including precise instructions for the work of school coordinators and scripts for test administrators to use in testing sessions.

Recruitment of schools and students

With the exception of private schools, the recruitment of schools required several steps. Beginning with the sampled schools, the first step entailed obtaining permission from the school district to approach the sampled school(s) in that district. If a district refused permission, then the district of the first substitute school was approached and the procedure was repeated. With permission from the district, the school(s) was contacted in a second step. If a sampled school refused to participate, the district of the first substitute was approached

and the permission procedure repeated. During most of the recruitment period sampled schools and substitute schools were being recruited concurrently. Each participating school was asked to nominate a school coordinator as the main point of contact for the study. The school coordinator worked with project staff to arrange logistics and liaise with staff, students and parents as necessary.

On the advice of the school, parental permission for students to participate was sought with one of three approaches to parents: a simple notification; a notification with a refusal form; and a notification with a consent form for parents to sign. In each approach, parents were informed that their students could opt out of participating.

Gifts to schools, school coordinators, and students

Schools, school coordinators, and students were provided with small gifts in appreciation for their willingness to participate. Schools were offered $200, school coordinators received $100, and students were given a clock-compass carabiner.

Test administration

Test administration in the United States was carried out by professional staff trained according to the international guidelines. School personnel were asked only to assist with listings of students, identifying space for testing in the school, and specifying any parental consent procedures needed for sampled students.

Calculator usage

Calculators were not permitted during the TIMSS 4th-grade assessment. However, the TIMSS policy on calculator use at the 8th grade was to give students the best opportunity to operate in settings that mirrored their classroom experiences. Calculators were permitted but not required for the 8th-grade assessment materials. In the United States, all students were allowed, but not required, to use a calculator.

Quality assurance

The International Study Center monitored compliance with the standardized procedures. National research coordinators were asked to nominate one or more persons unconnected with their national center, such as retired school teachers, to serve as quality control monitors for their country or education system. The International Study Center developed manuals for the monitors and briefed them in 2-day training sessions about TIMSS, the responsibilities of the national centers in conducting the study, and their own roles and responsibilities. Some 60 schools in the U.S. samples were visited by the monitors—30 schools in the 4th-grade sample, and 30 schools in the 8th-grade sample. These schools included those in both the national and state samples and were scattered geographically across the nation.

Scoring and scoring reliability

The TIMSS assessment items included both multiple-choice and constructed-response items. A scoring rubric (guide) was created for every constructed response item included in the TIMSS assessments. The rubrics were carefully written and reviewed by national research coordinators and other experts as part of the field test of items, and revised accordingly.

The national research coordinator in each country or education system was responsible for the scoring and coding of data for that participant, following established guidelines. The national research coordinator and, sometimes, additional staff attended scoring training sessions held by the International Study Center. The training sessions focused on the scoring rubrics and coding system employed in TIMSS. Participants in these training sessions were provided extensive practice in scoring example items over several days. Information on within-country agreement among coders was collected and documented by the International Study Center. Information on scoring and coding reliability was also used to calculate cross-country agreement among coders. Information on scoring reliability for constructed-response scoring in TIMSS 2011 is provided in the international report.

Data entry and cleaning

The national research coordinator from each country or education system was responsible for data entry. In the United States, Westat was contracted to collect data for TIMSS 2011 and entered the data into data files using a common international format. This format was specified in the *Data Entry Manager Manual* (IEA Data Processing Center 2010), which accompanied the IEA-supplied data-entry software (WinDEM) given to all participating countries to create data files. This software facilitated the checking and correction of data by providing various data consistency checks. The data were then sent to the IEA Data Processing Center (DPC) in Hamburg, Germany, for cleaning. The DPC checked that the international data structure was followed; checked the identification system within and between files; corrected single case problems manually; and applied standard cleaning procedures to questionnaire files. Results of the data cleaning process were documented by the DPC. This documentation was then sent to the national research coordinator along with any remaining questions about the data. The national research coordinator then provided the DPC with revisions to coding or solutions for anomalies. The DPC subsequently compiled background univariate statistics and preliminary test scores based on classical item analysis and item response theory (IRT). Detailed information on the entire data entry and cleaning process can be found in the *TIMSS and PIRLS Methods and Procedures* (Martin and Mullis 2011).

Weighting, scaling, and plausible values

Before the data were analyzed, responses from the groups of students assessed were assigned sampling weights to ensure that their representation in the TIMSS 2011 results matched their actual percentage of the school population in the grade assessed. With these sampling weights in place, the analyses of TIMSS 2011 data proceeded in two phases: scaling and estimation. During the scaling phase, IRT procedures were used to estimate the measurement characteristics of each assessment question. During the estimation phase, the results of the scaling were used to produce estimates of student achievement. Subsequent analyses related these achievement results to the background variables collected by TIMSS 2011.

Weighting

Responses from the groups of students were assigned sampling weights to adjust for over- or under-representation during the sampling of a particular group. The use of sampling weights is necessary for the computation of sound, nationally representative estimates. The weight assigned to a student's responses is the inverse of the probability that the student is selected for the sample. When responses are weighted, none are discarded, and each contributes to the results for the total number of students represented by the individual student assessed. Weighting also adjusts for various situations (such as school and student nonresponse) because data cannot be assumed to be randomly missing. The internationally defined weighting specifications for TIMSS require that each assessed student's sampling weight should be the product of (1) the inverse of the school's probability of selection, (2) an adjustment for school-level nonresponse, (3) the inverse of the classroom's probability of selection, and (4) an adjustment for student-level nonresponse.[19] All TIMSS 1995, 1999, 2003, 2007, and 2011 analyses are conducted using sampling weights. A detailed description of this process is provided in the *TIMSS and PIRLS Methods and Procedures* (Martin and Mullis 2011). For 2011, though the national and state samples share schools, the samples are not identical school samples and, thus, weights are estimated separately for the national and state samples.

Scaling

In TIMSS, the propensity of students to answer questions correctly was estimated with a two-parameter IRT model for dichotomous constructed response items, a three-parameter IRT model for multiple choice response items, and a generalized partial credit IRT model for polytomous constructed response items. The scale scores assigned to each student were estimated using a procedure described below in the "Plausible values" section, with input from the IRT results. With IRT, the difficulty of each item, or item category, is deduced using information about how likely it is for students to get some items correct (or to get a higher rating on a constructed response item) versus other items. Once the parameters of each item are determined, the ability of each student can be estimated even when different students have been administered different items. At this point in the estimation process achievement scores are expressed in a standardized logit scale that ranges from -4 to +4. In order to make the scores more meaningful and to facilitate their interpretation, the scores for the first year (1995) are transformed to a scale with a mean of 500 and a standard deviation of 100. Subsequent waves of assessment are linked to this metric (see below).

To make scores from the second (1999) wave of data comparable to the first (1995) wave, two steps had to be taken. First, the 1995 and 1999 data for countries and education systems that participated in both years were scaled together to estimate item parameters. Ability estimates for all students (those assessed in 1995 and those assessed in 1999) based on the new item parameters were then estimated. To put these jointly calibrated 1995 and 1999 scores on the 1995 metric, a linear transformation was applied such that the jointly calibrated 1995 scores have the same mean and standard deviation as the original 1995 scores. Such a transformation also preserves any differences in average scores between the 1995 and 1999 waves of assessment.

In order for scores resulting from subsequent waves of assessment (2003, 2007 and 2011) to be made comparable to 1995 scores (and to each other), the two steps above are applied sequentially for each pair of adjacent waves of data: two adjacent years of data are jointly scaled, then resulting ability estimates are linearly transformed so that the mean and standard deviation of the prior year is preserved. As a result, the transformed-2011 scores are comparable to all previous waves of the assessment and longitudinal comparisons between all waves of data are meaningful.

To facilitate the joint calibration of scores from adjacent years of assessment, common test items are included in successive administrations. This also enables the comparison of item parameters (difficulty and discrimination) across administrations. If item parameters change dramatically across administrations, they are dropped from the current assessment so that scales can be more accurately linked across years. In this way even if the average ability levels of students in countries and education systems participating in TIMSS changes over time, the scales still can be linked across administrations.

[19]These adjustments are for overall response rates and did not include any of the characteristics associated with differential nonresponse as identified in the nonresponse bias analyses reported above.

Plausible values

To keep student burden to a minimum, TIMSS administered a limited number of assessment items to each student—too few to produce accurate content-related scale scores for each student. To accommodate this situation, during the scaling process plausible values were estimated to characterize students participating in the assessment, given their background characteristics. Plausible values are imputed values and not test scores for individuals in the usual sense. In fact, they are biased estimates of the proficiencies of individual students. Plausible values do, however, provide unbiased estimates of population characteristics (e.g., means and variances of demographic subgroups).

Plausible values represent what the performance of an individual on the entire assessment might have been, had it been observed. They are estimated as random draws (usually five) from an empirically derived distribution of score values based on the student's observed responses to assessment items and on background variables. Each random draw from the distribution is considered a representative value from the distribution of potential scale scores for all students in the sample who have similar characteristics and identical patterns of item responses. Differences between plausible values drawn for a single individual quantify the degree of error (the width of the spread) in the underlying distribution of possible scale scores that could have caused the observed performances.

An accessible treatment of the derivation and use of plausible values can be found in Beaton and González (1995). A more technical treatment can be found in the *TIMSS and PIRLS Methods and Procedures* (Martin and Mullis 2011).

International benchmarks

International benchmarks for achievement were developed in an attempt to provide a concrete interpretation of what the scores on the TIMSS mathematics and science achievement scales mean (for example, what it means to have a scale score of 513 or 426). To describe student performance at various points along the TIMSS mathematics and science achievement scales, TIMSS uses scale anchoring to summarize and describe student achievement at four points on the mathematics and science scales—*Advanced* (625), *High* (550), *Intermediate* (475), and *Low* (400) international benchmarks. Scale anchoring involves selecting benchmarks (scale points) on the TIMSS achievement scales to be described in terms of student performance. Once benchmark scores have been chosen, items are identified that students are likely to score highly on. The content of these items describes what students at each benchmark level of achievement know and can do. To interpret the content of anchored items, these items are grouped by content area within benchmarks and reviewed by mathematics and science experts. These experts focus on the content of each item and describe the kind of mathematics or science knowledge demonstrated by students answering the item correctly. The experts then provide a summary description of performance at each anchor point leading to a content-referenced interpretation of the achievement results. Detailed information on the creation of the benchmarks is provided in the international TIMSS reports (Mullis et al. 2012; Martin et al. 2012).

Data limitations

As with any study, there are limitations to TIMSS 2011 that researchers should take into consideration. Estimates produced using data from TIMSS 2011 are subject to two types of error—nonsampling and sampling errors. Nonsampling errors can be due to errors made in collecting and processing data. Sampling errors can occur because the data were collected from a sample rather than a complete census of the population.

Nonsampling errors

Nonsampling error is a term used to describe variations in the estimates that may be caused by population coverage limitations, nonresponse bias, and measurement error, as well as data collection, processing, and reporting procedures. The sources of nonsampling errors are typically problems like unit and item nonresponse, differences in respondents' interpretations of the meaning of the survey questions, response differences related to the particular time the survey was conducted, and mistakes in data preparation.

Missing data. Five kinds of missing data were identified by separate missing data codes: omitted, uninterpretable, not administered, not applicable, and not reached. An item was considered omitted if the respondent was expected to answer the item but no response was given (e.g., no box was checked in the item which asked "Are you a girl or a boy?"). Items with invalid responses (e.g., multiple responses to a question calling for a single response) were coded as *uninterpretable*. The *not administered* code was used to identify items not administered to the student, teacher, or principal (e.g., those items excluded from the student's test booklet because of the BIB-spiraling of the items). An item was coded as *not applicable* when it is not logical that the respondent answer the question (e.g., when the opportunity to make the response is dependent on a filter question). Finally, items that are *not reached* were identified by a string of consecutive items without responses continuing through to the end of the assessment or questionnaire.

Missing background data on other than key variables are not included in the analyses for this report and are not imputed.[20] Item response rates for variables discussed in this report exceeded the NCES standard of 85 percent and so can be reported without notation. Of the three key variables identified in the TIMSS 2011 data for the United States—sex, race/ethnicity and the percentage of students eligible for free or reduced-price lunch (FRPL)—as table A-5 indicates, sex has no missing responses and race/ethnicity has minimal missing responses at some 2 percent. The FRPL variable has some 5 percent missing responses at 4th grade and 3 percent missing responses at 8th grade among the public schools in the sample and these were imputed by substituting values taken from the CCD for the schools in question. Note, however, that the CCD provides this information only for public schools. The comparable database for private schools (PSS) does not include data on participation in the FRPL program.

Sampling errors

Sampling errors arise when a sample of the population, rather than the whole population, is used to estimate some statistic. Different samples from the same population would likely produce somewhat different estimates of the statistic in question. This fact means that there is a degree of uncertainty associated with statistics estimated from a sample. This uncertainty is referred to as sampling variance and is usually expressed as the standard error of a statistic estimated from sample data. The approach used for calculating standard errors in TIMSS was jackknife repeated replication (JRR). Standard errors can be used as a measure for the precision expected from a particular sample. Standard errors for all of the reported estimates are included in appendix E.

Confidence intervals provide a way to make inferences about population statistics in a manner that reflects the sampling error associated with the statistic. Assuming a normal distribution, the population value of this statistic can be inferred to lie within the confidence interval in 95 out of 100 replications of the measurement on different samples drawn from the same population.

For example, the average mathematics score for the U.S. 8th-grade students was 509 in 2011, and this statistic had a standard error of 2.6. Therefore, it can be stated with 95 percent confidence that the actual average of U.S. 8th-grade students in 2011 was between 504 and 514 (1.96 x 2.6 = 5.1; confidence interval = 509 +/- 5.1).

Description of background variables

The international versions of the TIMSS 2011 student, teacher, and school questionnaires are available at http://timss.bc.edu. The U.S. versions of these questionnaires are available at http://nces.ed.gov/timss.

[20]Key variables include survey-specific items for which aggregate estimates are commonly published by NCES. They include, but are not restricted to, variables most commonly used in table row stubs. Key variables also include important analytic composites and other policy-relevant variables that are essential elements of the data collection. For example, the National Assessment of Educational Progress (NAEP) consistently uses sex, race/ethnicity, urbanicity, region, and school type (public/private) as key reporting variables.

Table A-5. Weighted response rates for unimputed variables for TIMSS, by grade: 2011

Variable	Source of information	Grade 4 U.S. response rate	Grade 4 Range of response rates in other countries	Grade 8 U.S. response rate	Grade 8 Range of response rates in other countries
Sex	Classroom tracking form	100	99.5 - 100	100	99.5 - 100
Race/ethnicity	Student questionnaire	98	†	99	†
Free or reduced-price lunch	School questionnaire	95	†	97	†

† Not applicable (U.S.-only variables).
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Race/ethnicity

Students' race/ethnicity was obtained through student responses to a two-part question. Students were asked first whether they were Hispanic or Latino, and then whether they were members of the following racial groups: American Indian or Alaska Native; Asian; Black or African American; Native Hawaiian or other Pacific Islander; or White. Multiple responses to the race classification question were allowed. Students who responded that they are Hispanic or Latino were categorized as Hispanic, regardless of their reported races. Results are shown separately for Blacks, Hispanics, Whites, Asians, and Multiracial as distinct groups. The small numbers of students indicating that they were American Indian or Alaska Native or Native Hawaiian or other Pacific Islander are included in the total but not reported separately.

Poverty level in public schools (percentage of students eligible for free or reduced-price lunch)

The poverty level in public schools was obtained from principals' responses to the school questionnaire. The question asked the principal to report, as of approximately the first of October 2010, the percentage of students at the school eligible to receive free or reduced-price lunch through the National School Lunch Program. The answers were grouped into five categories: less than 10 percent; 10 to 24.9 percent; 25 to 49.9 percent; 50 to 74.9 percent; and 75 percent or more. Analysis was limited to public schools only. Missing data on this variable were replaced with measures taken from the CCD. The effect of this replacement on the confidentiality of the data was examined as part of the confidentiality analyses described in the following section.

Confidentiality and disclosure limitations

In accord with NCES standard 4-2-6 (NCES Education Statistics 2002), confidentiality analyses for the United States were implemented to provide reasonable assurance that public-use data files issued by the IEA and NCES would not allow identification of individual U.S. schools or students when compared against publicly available data collections.

Disclosure limitations included the identification and masking of potential disclosure risks for TIMSS schools and adding an additional measure of uncertainty of school, teacher, and student identification through random swapping of a small number of data elements within the student, teacher, and school files.[21] These procedures were applied to the national and state samples.

Statistical procedures

Tests of significance

Comparisons made in the text of this report were tested for statistical significance. For example, in the commonly made comparison of education systems averages against the average of the United States, tests of statistical significance were used to establish whether or not the observed differences from the U.S. average were statistically significant. The estimation of the standard errors that is required in order to undertake the tests of significance is complicated by the complex sample and assessment designs, both of which generate error variance. Together they mandate a set of statistically complex procedures in order to estimate the correct standard errors. As a consequence, the estimated standard errors contain a sampling variance component estimated by the jackknife repeated replication (JRR) procedure; and, where the assessments are concerned, an additional imputation variance component arising from the assessment design. Details on the procedures used can be found in the *WesVar 5.0 User's Guide* (Westat 2007).

[21]The NCES standards describe such techniques as follows: perturbation disclosure limitation techniques directly alter the individual respondent's data for some variables, but preserve the level of detail in all variables included in the microdata file. The following are examples of perturbation techniques: blanking and imputing for randomly selected records; blurring (e.g., combining multiple records through some averaging process into a single record); adding random noise; and data swapping or switching (e.g., switching the sex variable from a predetermined pair of individuals) are all examples of perturbation techniques (National Center for Education Statistics 2002).

In almost all instances, the tests for significance used were standard *t* tests.[22] These fell into two categories according to the nature of the comparison being made: comparisons of independent samples and comparisons of nonindependent samples. Before describing the *t* tests used, some background on the two types of comparisons is provided below.

The variance of a difference is equal to the sum of the variances of the two initial variables minus two times the covariance between the two initial variables. A sampling distribution has the same characteristics as any distribution, except that units consist of sample estimates and not observations. Therefore,

$$\sigma^2_{(\hat{\mu}_X - \hat{\mu}_Y)} = \sigma^2_{(\hat{\mu}_X)} + \sigma^2_{(\hat{\mu}_Y)} - 2\mathrm{cov}(\hat{\mu}_X, \hat{\mu}_Y)$$

The sampling variance of a difference is equal to the sum of the two initial sampling variances minus two times the covariance between the two sampling distributions on the estimates.

If one wants to determine whether girls' performance differs from boys' performance, for example, then, as for all statistical analyses, a null hypothesis has to be tested. In this particular example, it consists of computing the difference between the boys' performance mean and the girls' performance mean (or the inverse). The null hypothesis is

$$H_0 : \hat{\mu}_{(boys)} - \hat{\mu}_{(girls)} = 0$$

To test this null hypothesis, the standard error on this difference is computed and then compared to the observed difference. The respective standard errors on the mean estimate for boys and girls can be easily computed.

The expected value of the covariance will be equal to 0 if the two sampled groups are independent. If the two groups are not independent, as is the case with girls and boys attending the same schools within an education system, or comparing an education system's mean with the international mean that includes that particular country, the expected value of the covariance might differ from 0.

In TIMSS, participating education systems' samples are independent. Therefore, for any comparison between two education systems, the expected value of the covariance will be equal to 0, and thus the standard error on the estimate is

$$\sigma_{(\hat{\theta}_i - \hat{\theta}_j)} = \sqrt{\sigma^2_{(\hat{\theta}_i)} + \sigma^2_{(\hat{\theta}_j)}}$$

with θ being a tested statistic.

Within a particular education system, any subsamples will be considered as independent only if the categorical variable used to define the subsamples was used as an explicit stratification variable.

Therefore, as for any computation of a standard error in TIMSS, replication methods using the supplied replicate weights are used to estimate the standard error on a difference. Use of the replicate weights implicitly incorporates the covariance between the two estimates into the estimate of the standard error on the difference.

Thus, in simple comparisons of independent averages, such as the U.S. average with other education systems' averages, the following formula was used to compute the *t* statistic:

$$t = \frac{(est_1 - est_2)}{\sqrt{(se_1)^2 + (se_2)^2}}$$

Est_1 and est_2 are the estimates being compared (e.g., average of education system A and the U.S. average), and se_1 and se_2 are the corresponding standard errors of these averages.

The second type of comparison used in this report occurred when comparing differences of nonsubset, nonindependent groups (e.g., when comparing the average scores of boys versus girls within the United States). In such comparisons, the following formula was used to compute the *t* statistic:

$$t = \frac{(est_{grp1} - est_{grp2})}{se(est_{grp1} - est_{grp2})}$$

Est_{grp1} and est_{grp2} are the nonindependent group estimates being compared. $Se(est_{grp1} - est_{grp2})$ is the standard error of the difference calculated using a JRR procedure, which accounts for any covariance between the estimates for the two nonindependent groups.

[22] Adjustments for multiple comparisons were not applied in any of the t-tests undertaken.

Appendix B: Example Items

After each administration of TIMSS, the IEA releases to the public somewhat less than half of the TIMSS items in order to illustrate the content of the assessment. The remaining items are kept secure so they can be used again in a future administration of TIMSS to measure trends in performance. This appendix contains sample mathematics and science items used in the U.S. administration of TIMSS 2011. These items have been selected from the set of released items to provide examples from each of the international benchmark levels, each of the content and cognitive domains, and each of the response types.

Exhibits B.1 and B.2 below provide a key to which items are examples of each of these dimensions, B.1 for mathematics and B.2 for science. Reading exhibit B.1, for example, one can see that two items illustrate the Number content domain at grade 4 (exhibits B.5 and B.6) but that each of these represents a different benchmark level, a different cognitive domain, and a different item response type. Each item is presented on a separate page in this appendix. For all multiple choice items, the test question and "response options" (possible answers) are reproduced on the page along with the "item key" (correct answer). For all constructed-response items, the "scoring rubric" (the criteria for scoring) is reproduced along with the test question. All item pages also include the percentage of students who received full credit for their answer in each participating country or other education system. Note that although most constructed response items were worth 1 point, some were worth 2 points with 1 point awarded for partial credit. In this appendix, if an example item was worth 2 points, only the percentages of students with responses awarded 2 points (full credit) are shown.

Exhibit B-1. Sample TIMSS 2011 mathematics items, by grade level, international benchmark level, content domain, cognitive domain, and item response type

Exhibit number, by grade	International benchmark level	Content domain	Cognitive domain	Item response type
Grade 4				
B.3	Low	Data Display	Applying	CR
B.4	Intermediate	Geometric Shapes and Measures	Knowing	MC
B.5	High	Number	Reasoning	MC
B.6	Advanced	Number	Applying	CR
Grade 8				
B.7	Low	Number	Knowing	CR
B.8	Intermediate	Algebra	Knowing	CR
B.9	High	Data and Chance	Reasoning	MC
B.10	Advanced	Geometry	Applying	MC

NOTE: CR indicates constructed-response item. MC indicates multiple choice item.

Exhibit B-2. Sample TIMSS 2011 science items, by grade level, international benchmark level, content domain, cognitive domain, and item response type

Exhibit number, by grade	International benchmark level	Content domain	Cognitive domain	Item response type
Grade 4				
B.11	Low	Life Science	Applying	MC
B.12	Intermediate	Physical Science	Knowing	MC
B.13	High	Life Science	Reasoning	CR
B.14	Advanced	Earth Science	Applying	CR
Grade 8				
B.15	Low	Chemistry	Applying	MC
B.16	Intermediate	Biology	Knowing	MC
B.17	High	Earth Science	Knowing	CR
B.18	Advanced	Physics	Reasoning	CR

NOTE: CR indicates constructed-response item. MC indicates multiple choice item.

Exhibit B-3. Example 4th-grade mathematics item: 2011

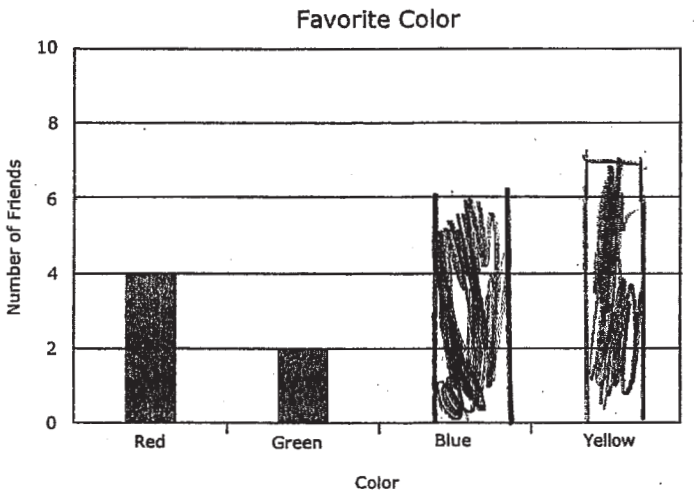
International Benchmark Level	Low
Content Domain	Data Display
Cognitive Domain	Applying

Favorite colors of Darin's friends

Darin asked his friends to name their favorite color. He collected the information in the table shown below.

Favorite Color	Number of Friends
Red	4
Green	2
Blue	6
Yellow	7

Then Darin started to draw a graph to show the information. Complete Darin's graph.



M031133

Education system	Percent full credit
International average	73
Korea, Rep. of	97
Singapore[1]	95
Hong Kong-CHN[1]	95
Japan	93
Northern Ireland-GBR[2]	92
Netherlands[2]	91
England-GBR	89
Finland	88
Germany	88
Lithuania[1,3]	87
Ireland	87
Chinese Taipei-CHN	87
Belgium (Flemish)-BEL	86
Australia	84
Portugal	84
Denmark[1]	84
Sweden	83
Malta	83
Hungary	83
Russian Federation	81
New Zealand	81
Austria	80
Slovenia	80
Thailand	78
United States[1]	**78**
Spain	78
Slovak Republic	77
Czech Republic	77
Italy	77
Bahrain	75
Croatia[1]	74
Norway[4]	74
Turkey	73
Kazakhstan[1]	73
Poland	73
Qatar[1]	70
Chile	69
United Arab Emirates	68
Serbia[1]	67
Romania	62
Saudi Arabia	60
Oman[5]	57
Georgia[3,6]	56
Kuwait[3,7]	**55**
Iran, Islamic Rep. of	54
Azerbaijan[1,6]	47
Armenia	41
Tunisia[5]	24
Morocco[7]	23
Yemen[7]	13

Benchmarking education systems	Percent full credit
Quebec-CAN	89
Ontario-CAN	87
North Carolina-USA[1,3]	82
Alberta-CAN[1]	81
Florida-USA[3,8]	80
Dubai-UAE	75
Abu Dhabi-UAE	62

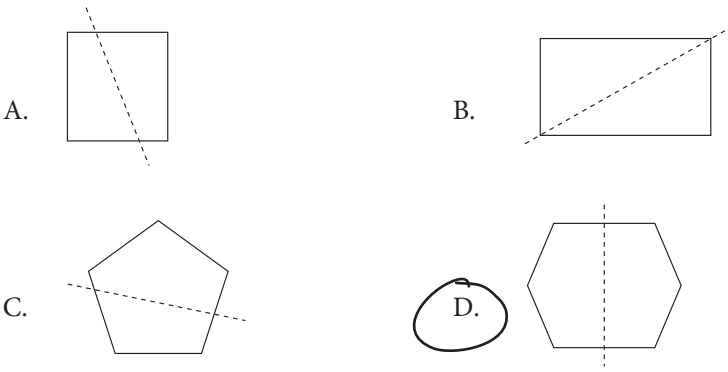
[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Met guidelines for sample participation rates only after replacement schools were included.
[3]National Target Population does not include all of the International Target Population (see appendix A).
[4]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[5]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[6]Exclusion rates for Azerbaijan and Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[7]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
[8]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Exhibit B-4. Example 4th-grade mathematics item: 2011

International Benchmark Level	Intermediate
Content Domain	Geometric Shapes and Measures
Cognitive Domain	Knowing

Which dotted line is a line of symmetry?

In which of the following figures is the dotted line a line of symmetry?



M031093

Education system	Percent full credit
International average	54
Singapore[1]	91
Hong Kong-CHN[1]	81
United States[1]	**80**
Australia	76
Korea, Rep. of	75
Northern Ireland-GBR[2]	74
England-GBR	71
Malta	70
Ireland	66
Turkey	65
Denmark[1]	65
Saudi Arabia	65
Germany	64
Iran, Islamic Rep. of	64
Portugal	63
Italy	63
Czech Republic	63
Hungary	61
Lithuania[1,3]	61
Kazakhstan[1]	59
Russian Federation	58
Kuwait[3,4]	58
Bahrain	57
Oman[5]	54
United Arab Emirates	54
Chile	53
New Zealand	53
Thailand	53
Norway[6]	52
Azerbaijan[1,7]	52
Romania	50
Qatar[1]	50
Slovenia	49
Austria	49
Belgium (Flemish)-BEL	49
Finland	49
Spain	48
Georgia[3,7]	44
Serbia[1]	43
Chinese Taipei-CHN	43
Poland	40
Slovak Republic	40
Armenia	39
Morocco[4]	**37**
Sweden	32
Netherlands[2]	30
Croatia[1]	30
Japan	30
Yemen[4]	27
Tunisia[5]	22

Benchmarking education systems	Percent full credit
Florida-USA[3,8]	87
North Carolina-USA[1,3]	86
Ontario-CAN	76
Alberta-CAN[1]	62
Quebec-CAN	62
Abu Dhabi-UAE	54
Dubai-UAE	52

[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Met guidelines for sample participation rates only after replacement schools were included.
[3]National Target Population does not include all of the International Target Population (see appendix A).
[4]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
[5]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[6]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[7]Exclusion rates for Azerbaijan and Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[8]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Exhibit B-5. Example 4th-grade mathematics item: 2011

International Benchmark Level	High
Content Domain	Number
Cognitive Domain	Reasoning

Distance between towns using map

The scale on a map indicates that 1 centimeter on the map represents 4 kilometers on the land. The distance between the two towns on the map is 8 centimeters. How many kilometers apart are the two towns?

A. 2

B. 8

C. 16

D. 32

M031185

Education system	Percent full credit
International average	54
Korea, Rep. of	84
Chinese Taipei-CHN	84
Russian Federation	80
Lithuania[1,2]	75
Japan	74
Finland	71
Serbia[2]	71
Singapore[2]	69
Netherlands[3]	69
Kazakhstan[2]	69
Czech Republic	67
Azerbaijan[2,4]	66
Croatia[2]	64
Denmark[2]	63
Slovenia	63
Northern Ireland-GBR[3]	62
Slovak Republic	61
Germany	60
Hungary	59
United States[2]	**59**
Portugal	57
Romania	57
Sweden	57
Austria	57
Poland	57
Norway[5]	56
Iran, Islamic Rep. of	55
England-GBR	54
Italy	54
Belgium (Flemish)-BEL	53
Ireland	50
Turkey	48
Georgia[1,4]	47
Australia	46
Spain	46
Armenia	45
New Zealand	42
Bahrain	40
Chile	39
Thailand	39
Saudi Arabia	39
United Arab Emirates	37
Qatar[2]	32
Malta	32
Morocco[6]	31
Oman[7]	31
Tunisia[7]	31
Yemen[6]	29
Kuwait[1,6]	23
Hong Kong-CHN[2]	—

Benchmarking education systems	Percent full credit
North Carolina-USA[1,2]	62
Florida-USA[1,8]	62
Quebec-CAN	57
Ontario-CAN	50
Alberta-CAN[2]	43
Dubai-UAE	42
Abu Dhabi-UAE	33

— Not available.
[1]National Target Population does not include all of the International Target Population (see appendix A).
[2]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[3]Met guidelines for sample participation rates only after replacement schools were included.
[4]Exclusion rates for Azerbaijan and Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[5]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
[7]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[8]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Exhibit B-6. Example 4th-grade mathematics item: 2011

International Benchmark Level	Advanced
Content Domain	Number
Cognitive Domain	Applying

Recipe for 3 people

Ingredients	
Eggs	4
Flour	8 cups
Milk	$\frac{1}{2}$ cup

The above ingredients are used to make a recipe for 6 people. Sam wants to make this recipe for only 3 people.

Complete the table below to show what Sam needs to make the recipe for 3 people. The number of eggs he needs is shown.

Ingredients	
Eggs	2
Flour	4 cups
Milk	$\frac{1}{4}$ cup

M031183

Education system	Percent full credit
International Avg.	23
Northern Ireland-GBR[1]	63
England-GBR	55
Ireland	54
Singapore[2]	50
Germany	46
Netherlands[1]	41
New Zealand	39
Belgium (Flemish)-BEL	38
Denmark[2]	37
Australia	37
Hong Kong-CHN[2]	34
United States[2]	**33**
Malta	31
Finland	31
Chinese Taipei-CHN	31
Portugal	28
Korea, Rep. of	28
Serbia[2]	27
Lithuania[2,3]	24
Japan	23
Austria	23
Kazakhstan[2]	22
Spain	22
Romania	22
Qatar[2]	21
Bahrain	20
Azerbaijan[2,4]	20
Russian Federation	20
United Arab Emirates	20
Hungary	18
Saudi Arabia	18
Slovenia	17
Poland	16
Norway[5]	15
Sweden	15
Armenia	15
Chile	15
Italy	14
Georgia[3,4]	13
Oman[6]	11
Czech Republic	10
Slovak Republic	10
Kuwait[3,7]	8
Turkey	8
Thailand	5
Tunisia[6]	5
Morocco[7]	4
Croatia[2]	3
Yemen[7]	3
Iran, Islamic Rep. of	3

Benchmarking education systems	Percent full credit
North Carolina-USA[2,3]	32
Florida-USA[3,8]	31
Dubai-UAE	29
Quebec-CAN	22
Alberta-CAN[2]	22
Ontario-CAN	22
Abu Dhabi-UAE	15

[1]Met guidelines for sample participation rates only after replacement schools were included.
[2]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[3]National Target Population does not include all of the International Target Population (see appendix A).
[4]Exclusion rates for Azerbaijan and Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[5]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[7]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
[8]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Exhibit B-7. Example 8th-grade mathematics item: 2011

International Benchmark Level	Low
Content Domain	Number
Cognitive Domain	Knowing

Add 42.65 to 5.748

M052231

42.65 + 5.748 =

Answer: 48.398

Education system	Percent full credit
International Avg.	72
Singapore[1]	94
Malaysia	91
Hong Kong-CHN	91
Kazakhstan	90
Lithuania[1]	90
Russian Federation[1]	90
Chinese Taipei-CHN	89
United States[1]	**89**
Hungary	88
Italy	88
Korea, Rep. of	87
Slovenia	85
Armenia	84
Tunisia	82
Israel[2]	82
Australia	82
Norway	81
Lebanon	81
Japan	81
Ukraine	80
United Arab Emirates	79
Sweden	79
England-GBR[3]	79
Finland	79
Morocco[4]	72
Qatar[5]	72
New Zealand	70
Romania	69
Saudi Arabia[5]	65
Macedonia, Rep. of[5]	65
Georgia[6,7]	64
Thailand	64
Chile	58
Indonesia[5]	57
Palestinian Nat'l Auth.[5]	56
Oman[5]	49
Turkey	48
Bahrain[5]	43
Iran, Islamic Rep. of[5]	42
Jordan[5]	36
Ghana[4]	36
Syrian Arab Republic[5]	31

Benchmarking education systems	Percent full credit
Massachusetts-USA[1,6]	95
Minnesota-USA[6]	93
Florida-USA[1,6]	93
Alabama-USA[6]	92
Connecticut-USA[1,6]	91
Indiana-USA[1,6]	90
North Carolina-USA[2,6]	90
Quebec-CAN	90
California-USA[1,6]	89
Alberta-CAN[1]	86
Ontario-CAN[1]	85
Colorado-USA[6]	82
Abu Dhabi-UAE	81
Dubai-UAE	80

[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
[3]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[4]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
[5]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[6]National Target Population does not include all of the International Target Population (see appendix A).
[7]Exclusion rates for Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Exhibit B-8. Example 8th-grade mathematics item: 2011

International Benchmark Level	Intermediate
Content Domain	Algebra
Cognitive Domain	Knowing

Next term in the pattern

$$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}$$

A. What is the next term in this pattern?

Answer: $\frac{6}{7}$

M042198A

Education system	Percent full credit
International Avg.	70
Japan	94
Singapore[1]	94
Korea, Rep. of	92
Hong Kong-CHN	92
Chinese Taipei-CHN	92
Finland	91
Australia	91
Slovenia	90
Hungary	90
New Zealand	89
England-GBR[2]	89
Sweden	88
Norway	88
United States[1]	**87**
Lithuania[3]	86
Israel[4]	85
Italy	85
Ukraine	85
Russian Federation[1]	82
Romania	81
Kazakhstan	78
Turkey	77
Tunisia	76
United Arab Emirates	70
Iran, Islamic Rep. of[5]	68
Thailand	68
Armenia	66
Macedonia, Rep. of[5]	66
Bahrain[5]	62
Qatar[5]	56
Georgia[3,6]	53
Malaysia	53
Jordan[5]	45
Syrian Arab Republic[5]	44
Lebanon	43
Ghana[7]	38
Oman[5]	37
Palestinian Nat'l Auth.[5]	36
Saudi Arabia[5]	33
Indonesia[5]	32
Chile	32
Morocco[7]	15

Benchmarking education systems	Percent full credit
Massachusetts-USA[1,3]	94
Quebec-CAN	92
Ontario-CAN[1]	92
Indiana-USA[1,3]	91
Alberta-CAN[1]	90
Colorado-USA[3]	90
Minnesota-USA[3]	90
North Carolina-USA[3,4]	89
Connecticut-USA[1,3]	88
Florida-USA[1,3]	87
Alabama-USA[3]	87
California-USA[1,3]	86
Dubai-UAE	79
Abu Dhabi-UAE	68

[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[3]National Target Population does not include all of the International Target Population (see appendix A).
[4]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
[5]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[6]Exclusion rates for Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[7]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Exhibit B-9. Example 8th-grade mathematics item: 2011

International Benchmark Level	High
Content Domain	Data and Chance
Cognitive Domain	Reasoning

Probability that the marble is red

There are 10 marbles in a bag: 5 red, and 5 blue.

Sue draws a marble from the bag at random. The marble is red.

She puts the marble back into the bag.

What is the probability that the next marble she draws at random is red?

A. $\frac{1}{2}$

B. $\frac{4}{10}$

C. $\frac{1}{5}$

D. $\frac{1}{10}$

M052429

Education system	Percent full credit
International Avg.	45
Singapore[1]	78
Korea, Rep. of	76
Australia	75
England-GBR[2]	72
Japan	70
Chinese Taipei-CHN	70
New Zealand	68
Slovenia	67
United States[1]	**66**
Finland	66
Sweden	65
Norway	64
Hong Kong-CHN	64
Israel[3]	59
Hungary	51
Lithuania[4]	51
Russian Federation[1]	47
Turkey	44
Italy	42
Macedonia, Rep. of[5]	41
Romania	39
Armenia	39
Palestinian Nat'l Auth.[5]	38
Ukraine	38
Kazakhstan	38
Thailand	37
United Arab Emirates	36
Indonesia[5]	35
Iran, Islamic Rep. of[5]	35
Saudi Arabia[5]	35
Qatar[5]	34
Georgia[4,6]	33
Chile	31
Ghana[7]	28
Malaysia	28
Jordan[5]	26
Bahrain[5]	24
Oman[5]	23
Lebanon	21
Tunisia	20
Syrian Arab Republic[5]	18
Morocco[7]	16

Benchmarking education systems	Percent full credit
Massachusetts-USA[1,4]	79
Minnesota-USA[4]	78
Connecticut-USA[1,4]	78
Colorado-USA[4]	76
Indiana-USA[1,4]	76
North Carolina-USA[3,4]	74
Alberta-CAN[1]	73
Ontario-CAN[1]	73
Quebec-CAN	72
Florida-USA[1,4]	60
California-USA[1,4]	60
Alabama-USA[4]	52
Dubai-UAE	45
Abu Dhabi-UAE	34

[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[3]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
[4]National Target Population does not include all of the International Target Population (see appendix A).
[5]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[6]Exclusion rates for Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[7]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Exhibit B-10. Example 8th-grade mathematics item: 2011

International Benchmark Level	Advanced
Content Domain	Geometry
Cognitive Domain	Applying

Degrees minute hand of clock turns

How many degrees does a minute hand of a clock turn through from 6:20 a.m. to 8:00 a.m. on the same day?

A. 680°
B. 600°
C. 540°
D. 420°

M032331

Education system	Percent full credit
International Avg.	29
Korea, Rep. of	63
Japan	58
Singapore[1]	57
Chinese Taipei-CHN	56
Hong Kong-CHN	51
Finland	32
Sweden	31
England-GBR[2]	31
Slovenia	31
Morocco[3]	30
Hungary	30
Syrian Arab Republic[4]	30
Palestinian Nat'l Auth.[4]	29
Russian Federation[1]	29
Saudi Arabia[4]	28
Macedonia, Rep. of[4]	28
Turkey	27
Israel[5]	27
Australia	27
New Zealand	26
Iran, Islamic Rep. of[4]	26
Tunisia	26
Malaysia	26
Ukraine	25
Armenia	25
Italy	25
Jordan[4]	25
Lebanon	24
Bahrain[4]	24
Romania	23
Norway	23
Kazakhstan	23
United Arab Emirates	23
United States[1]	**22**
Qatar[4]	22
Oman[4]	21
Lithuania[6]	21
Ghana[3]	21
Georgia[6,7]	19
Indonesia[4]	19
Thailand	16
Chile	—

Benchmarking education systems	Percent full credit
Quebec-CAN	30
Minnesota-USA[6]	29
Ontario-CAN[1]	26
North Carolina-USA[5,6]	26
Massachusetts-USA[1,6]	25
Dubai-UAE	24
Connecticut-USA[1,6]	24
Abu Dhabi-UAE	23
Indiana-USA[1,6]	19
Alberta-CAN[1]	19
Alabama-USA[6]	18
Colorado-USA[6]	18
Florida-USA[1,6]	18
California-USA[1,6]	17

— Not available.
[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[3]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
[4]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[5]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
[6]National Target Population does not include all of the International Target Population (see appendix A).
[7]Exclusion rates for Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Exhibit B-11. Example 4th-grade science item: 2011

International Benchmark Level	Low
Content Domain	Life Science
Cognitive Domain	Applying

Birds/bats/butterflies share

S031230

What do birds, bats and butterflies have in common?

A. feathers

B. hair

C. internal skeleton

D. wings

Education system	Percent full credit
International Avg.	83
Korea, Rep. of	99
United States[1]	**96**
Croatia[1]	95
Singapore[1]	95
Finland	95
Sweden	95
Ireland	95
Austria	94
England-GBR	94
Norway[2]	93
Germany	93
New Zealand	93
Portugal	92
Russian Federation	92
Australia	92
Slovenia	91
Netherlands[3]	91
Northern Ireland-GBR[3]	91
Denmark[1]	91
Serbia[1]	91
Czech Republic	90
Poland	90
Slovak Republic	89
Italy	89
Lithuania[1,4]	89
Belgium (Flemish)-BEL	88
Spain	87
Japan	87
Thailand	86
Georgia[4,5]	86
Hungary	84
Chile	84
Armenia	83
Chinese Taipei-CHN	83
Romania	83
Malta	82
Hong Kong-CHN[1]	79
Kazakhstan[1]	79
Turkey	79
Bahrain	75
Azerbaijan[1,5]	75
United Arab Emirates	74
Saudi Arabia	70
Iran, Islamic Rep. of	62
Qatar[1]	62
Tunisia[6]	61
Oman	61
Kuwait[4,6]	54
Morocco[7]	47
Yemen[7]	31

Benchmarking education systems	Percent full credit
Florida-USA[4,8]	97
Alberta-CAN[1]	96
North Carolina-USA[1,4]	95
Ontario-CAN	93
Quebec-CAN	92
Dubai-UAE	79
Abu Dhabi-UAE	70

[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[3]Met guidelines for sample participation rates only after replacement schools were included.
[4]National Target Population does not include all of the International Target Population (see appendix A).
[5]Exclusion rates for Azerbaijan and Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[7]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
[8]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Exhibit B-12. Example 4th-grade science item: 2011

International Benchmark Level	Intermediate
Content Domain	Physical Science
Cognitive Domain	Knowing

Temperature of ice, steam, water

S051086

Water, ice, and steam all have different temperatures.

What is the order from coldest to hottest?

A. ice, water, steam
B. ice, steam, water
C. steam, ice, water
D. steam, water, ice

Education system	Percent full credit
International Avg.	73
United States[1]	**90**
Netherlands[2]	88
Singapore[1]	87
Croatia[1]	87
Czech Republic	86
Hong Kong-CHN[1]	86
Italy	84
Russian Federation	84
Serbia[1]	84
Belgium (Flemish)-BEL	84
Australia	83
Slovak Republic	83
Denmark[1]	83
Finland	82
Spain	81
Hungary	81
Slovenia	80
Chile	80
England-GBR	80
Chinese Taipei-CHN	80
Korea, Rep. of	80
Austria	79
Northern Ireland-GBR[2]	79
Germany	79
Sweden	79
New Zealand	78
Ireland	76
Norway[3]	75
Kazakhstan[1]	73
Japan	72
Turkey	71
Romania	71
Bahrain	71
Lithuania[1,4]	70
Malta	70
United Arab Emirates	69
Saudi Arabia	67
Azerbaijan[1,5]	65
Poland	63
Georgia[4,5]	62
Iran, Islamic Rep. of	61
Qatar[1]	61
Armenia	59
Oman	56
Kuwait[4,6]	55
Thailand	55
Tunisia[6]	41
Morocco[7]	37
Yemen[7]	29
Portugal	—

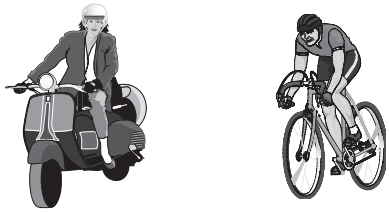
Benchmarking education systems	Percent full credit
Florida-USA[4,8]	94
North Carolina-USA[1,4]	90
Alberta-CAN[1]	86
Ontario-CAN	85
Dubai-UAE	73
Quebec-CAN	72
Abu Dhabi-UAE	70

— Not available.
[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Met guidelines for sample participation rates only after replacement schools were included.
[3]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[4]National Target Population does not include all of the International Target Population (see appendix A).
[5]Exclusion rates for Azerbaijan and Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[7]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
[8]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Exhibit B-13. Example 4th-grade science item: 2011

International Benchmark Level	High
Content Domain	Life Science
Cognitive Domain	Reasoning

Better way to travel around town



The pictures above show two ways of traveling around town.

A. Which way of traveling is better for the environment?

(Check one box.)

☑ Bicycle

☐ Motorbike

B. Explain your answer.

Because the motorcicle
send oil up to the atmosphere
and the ozone layer.

S041039

Education system	Percent full credit
International Avg.	48
Korea, Rep. of	83
Croatia[1]	78
Portugal	75
Slovenia	73
Finland	70
Italy	70
Sweden	68
Hungary	68
Russian Federation	67
Chinese Taipei-CHN	67
Spain	64
Czech Republic	64
Chile	63
Serbia[1]	62
Germany	62
Iran, Islamic Rep. of	61
Slovak Republic	60
Austria	60
Singapore[1]	54
Poland	54
Netherlands[2]	53
Belgium (Flemish)-BEL	53
Romania	51
Lithuania[1,3]	50
Norway[4]	49
England-GBR	47
Hong Kong-CHN[1]	45
Japan	45
Denmark[1]	44
United States[1]	**43**
Northern Ireland-GBR[2]	43
New Zealand	42
Australia	42
Ireland	41
Kazakhstan[1]	40
Bahrain	39
Turkey	38
Thailand	36
Tunisia[5]	35
United Arab Emirates	31
Malta	30
Qatar[1]	29
Armenia	29
Saudi Arabia	28
Georgia[3,6]	28
Morocco[7]	24
Kuwait[3,5]	20
Oman	18
Azerbaijan[1,6]	12
Yemen[7]	4

Benchmarking education systems	Percent full credit
Alberta-CAN[1]	54
Ontario-CAN	51
Florida-USA[3,8]	46
Quebec-CAN	45
Dubai-UAE	40
North Carolina-USA[1,3]	36
Abu Dhabi-UAE	31

[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Met guidelines for sample participation rates only after replacement schools were included.
[3]National Target Population does not include all of the International Target Population (see appendix A).
[4]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[5]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[6]Exclusion rates for Azerbaijan and Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[7]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
[8]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Exhibit B-14. Example 4th-grade science item: 2011

International Benchmark Level	Advanced
Content Domain	Earth Science
Cognitive Domain	Applying

Disadvantage to farming by a river

The picture below shows a river flowing across a plain.

Farming is carried out on the plain and near the river.

There are advantages and disadvantages to farming along a river.

B. Describe one **disadvantage**.

The river might be dirty

S041201B

[1]Met guidelines for sample participation rates only after replacement schools were included.
[2]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[3]National Target Population does not include all of the International Target Population (see appendix A). not covered and no official statistics were available.
[4]Exclusion rates for Azerbaijan and Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[5]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[7]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 25 percent.
[8]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Education system	Percent full credit
International Avg.	34
Korea, Rep. of	64
Czech Republic	60
Italy	55
Finland	55
Slovak Republic	55
Romania	53
Thailand	52
Chinese Taipei-CHN	52
Netherlands[1]	52
Slovenia	51
Singapore[2]	49
Austria	47
Ireland	46
Germany	46
Hong Kong-CHN[2]	45
Denmark[2]	44
Poland	44
Portugal	44
Hungary	43
Northern Ireland-GBR[1]	43
England-GBR	43
Russian Federation	42
Belgium (Flemish)-BEL	39
New Zealand	38
Australia	36
United States[2]	**35**
Lithuania[2,3]	34
Sweden	33
Turkey	32
Georgia[3,4]	29
Japan	26
Kazakhstan[2]	25
Azerbaijan[2,4]	25
Norway[5]	25
Spain	24
Serbia[2]	24
Chile	23
Croatia[2]	22
Iran, Islamic Rep. of	22
Malta	16
Bahrain	15
Armenia	15
United Arab Emirates	14
Qatar[2]	13
Saudi Arabia	11
Oman	7
Tunisia[6]	7
Morocco[7]	6
Kuwait[3,6]	5
Yemen[7]	1

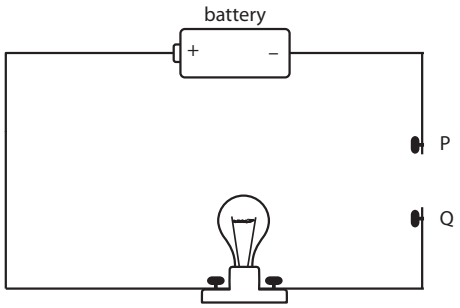
Benchmarking education systems	Percent full credit
Alberta-CAN[2]	42
Ontario-CAN	36
Quebec-CAN	35
North Carolina-USA[2,3]	25
Florida-USA[3,8]	24
Dubai-UAE	21
Abu Dhabi-UAE	9

Exhibit B-15. Example 8th-grade science item: 2011

International Benchmark Level	Low
Content Domain	Chemistry
Cognitive Domain	Applying

Which rod causes the bulb to light?

Rods made of different materials are connected between points P and Q in the circuit diagram shown below.



Which rod would cause the bulb to light?

A. copper rod
B. wood rod
C. glass rod
D. plastic rod

S042063

Education system	Percent full credit
International Avg.	88
Russian Federation[1]	97
Hong Kong-CHN	96
Lithuania[2]	96
Singapore[1]	96
Israel[3]	95
Slovenia	95
England-GBR[4]	95
Finland	94
Chinese Taipei-CHN	94
Japan	94
Chile	94
Thailand	93
Sweden	93
Indonesia	92
New Zealand	92
Turkey	92
Iran, Islamic Rep. of	91
Italy	91
Morocco	91
United States[1]	**90**
Australia	89
Tunisia	88
Korea, Rep. of	88
Jordan	88
Palestinian Nat'l Auth.	87
Norway	87
Romania	87
Syrian Arab Republic	87
Hungary	87
Ukraine	86
United Arab Emirates	84
Malaysia	84
Bahrain	83
Macedonia, Rep. of	83
Qatar	80
Saudi Arabia	80
Kazakhstan	80
Georgia[2,5]	80
Armenia	79
Lebanon	78
Oman	73
Ghana[6]	69

Benchmarking education systems	Percent full credit
Alberta-CAN[1]	95
Indiana-USA[1,2]	95
Minnesota-USA[2]	93
Massachusetts-USA[1,2]	93
North Carolina-USA[2,3]	93
Connecticut-USA[1,2]	92
Florida-USA[1,2]	90
Ontario-CAN[1]	90
Quebec-CAN	90
Colorado-USA[2]	90
Dubai-UAE	90
Alabama-USA[2]	87
California-USA[1,2]	85
Abu Dhabi-UAE	83

[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]National Target Population does not include all of the International Target Population (see appendix A).
[3]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
[4]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[5]Exclusion rates for Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Exhibit B-16. Example 8th-grade science item: 2011

International Benchmark Level	Intermediate
Content Domain	Biology
Cognitive Domain	Knowing

Cells that destroy bacteria

Bacteria that enter the body are destroyed by which type of cells?

A. white blood cells
B. red blood cells
C. kidney cells
D. lung cells

S032465

Education system	Percent full credit
International Avg.	61
Chinese Taipei-CHN	86
Singapore[1]	84
Korea, Rep. of	80
Italy	78
Japan	77
United States[1]	**76**
Sweden	74
Thailand	73
England-GBR[2]	71
Iran, Islamic Rep. of	71
Australia	70
Israel[3]	70
Lithuania[4]	68
Lebanon	68
Tunisia	68
Finland	68
Saudi Arabia	67
Kazakhstan	67
Hong Kong-CHN	66
Indonesia	66
Hungary	64
New Zealand	62
Romania	60
Macedonia, Rep. of	60
Syrian Arab Republic	60
Russian Federation[1]	59
Qatar	59
Bahrain	58
United Arab Emirates	57
Armenia	55
Malaysia	54
Norway	54
Palestinian Nat'l Auth.	52
Chile	48
Jordan	48
Oman	43
Ukraine	42
Ghana[5]	40
Turkey	37
Georgia[4,6]	35
Morocco	31
Slovenia	30

Benchmarking education systems	Percent full credit
Indiana-USA[1,4]	85
Minnesota-USA[4]	84
Massachusetts-USA[1,4]	79
Connecticut-USA[1,4]	79
North Carolina-USA[3,4]	79
Florida-USA[1,4]	78
Alberta-CAN[1]	77
Ontario-CAN[1]	77
Colorado-USA[4]	76
Alabama-USA[4]	74
Dubai-UAE	70
California-USA[1,4]	69
Quebec-CAN	60
Abu Dhabi-UAE	56

[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[3]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
[4]National Target Population does not include all of the International Target Population (see appendix A).
[5]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
[6]Exclusion rates for Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Exhibit B-17. Example 8th-grade science item: 2011

International Benchmark Level	High
Content Domain	Earth Science
Cognitive Domain	Knowing

Volcanic eruption effects

State one way that a volcanic eruption can affect the environment.

The hot magma can melt any thing away

S032126

Education system	Percent full credit
International Avg.	48
Singapore[1]	81
Slovenia	78
Sweden	76
Finland	71
New Zealand	70
Lithuania[2]	70
England-GBR[3]	67
Russian Federation[1]	63
Japan	63
Australia	63
United States[1]	**62**
Chile	62
Korea, Rep. of	58
Kazakhstan	58
Romania	57
Chinese Taipei-CHN	55
Hong Kong-CHN	54
Hungary	54
Norway	49
Turkey	49
Israel[4]	49
Ukraine	49
Thailand	47
Indonesia	45
Saudi Arabia	45
United Arab Emirates	42
Italy	41
Iran, Islamic Rep. of	37
Georgia[2,5]	34
Armenia	32
Bahrain	32
Jordan	32
Qatar	32
Malaysia	32
Macedonia, Rep. of	31
Palestinian Nat'l Auth.	31
Lebanon	28
Tunisia	28
Syrian Arab Republic	27
Oman	26
Morocco	19
Ghana[6]	9

Benchmarking education systems	Percent full credit
Massachusetts-USA[1,2]	70
Ontario-CAN[1]	67
Colorado-USA[2]	67
Connecticut-USA[1,2]	65
Minnesota-USA[2]	65
Florida-USA[1,2]	64
Alberta-CAN[1]	61
Indiana-USA[1,2]	59
California-USA[1,2]	58
North Carolina-USA[2,4]	57
Quebec-CAN	56
Dubai-UAE	51
Alabama-USA[2]	46
Abu Dhabi-UAE	39

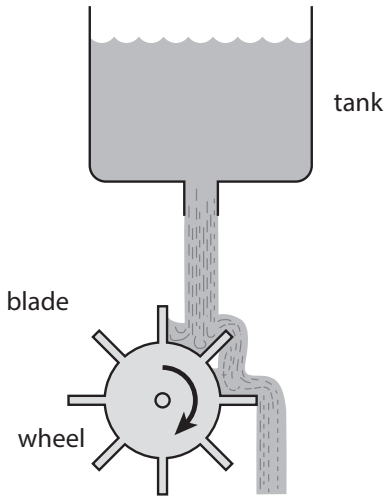
[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]National Target Population does not include all of the International Target Population (see appendix A).
[3]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[4]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
[5]Exclusion rates for Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Exhibit B-18. Example 8th-grade science item: 2011

International Benchmark Level	Advanced
Content Domain	Physics
Cognitive Domain	Reasoning

Water wheel: Faster rotation

The diagram shows water flowing from a tank and rotating a wheel.



C. Write one change to the system that will make the wheel rotate faster.

more water coming out at one; a bigger opening

S052165C

Education system	Percent full credit
International average	27
Singapore[1]	67
Japan	58
Hong Kong-CHN	46
Korea, Rep. of	44
Israel[2]	44
Chinese Taipei-CHN	44
England-GBR[3]	42
Finland	41
Iran, Islamic Rep. of	40
Turkey	37
Russian Federation[1]	37
Australia	36
Slovenia	35
Hungary	34
Norway	31
Ukraine	31
Lithuania[4]	31
New Zealand	29
United States[1]	**28**
Sweden	26
Syrian Arab Republic	25
Romania	25
Italy	23
Oman	23
Kazakhstan	20
Tunisia	20
Palestinian Nat'l Auth.	20
Bahrain	18
Jordan	18
United Arab Emirates	18
Saudi Arabia	17
Macedonia, Rep. of	17
Qatar	17
Malaysia	16
Armenia	16
Georgia[4,5]	13
Chile	12
Lebanon	11
Thailand	10
Indonesia	9
Morocco	5
Ghana[6]	3

Benchmarking education systems	Percent full credit
Massachusetts-USA[1,4]	37
Minnesota-USA[4]	35
Alberta-CAN[1]	35
Colorado-USA[4]	33
Connecticut-USA[1,4]	33
Ontario-CAN[1]	32
Quebec-CAN	31
Indiana-USA[1,4]	25
Florida-USA[1,4]	25
Dubai-UAE	24
North Carolina-USA[2,4]	23
Alabama-USA[4]	17
California-USA[1,4]	17
Abu Dhabi-UAE	17

[1]National Defined Population covers 90 to 95 percent of National Target Population (see appendix A).
[2]National Defined Population covers less than 90 percent, but at least 77 percent, of National Target Population (see appendix A).
[3]Nearly satisfied guidelines for sample participation rates after replacement schools were included.
[4]National Target Population does not include all of the International Target Population (see appendix A).
[5]Exclusion rates for Georgia are slightly underestimated as some conflict zones were not covered and no official statistics were available.
[6]The TIMSS International Study Center has reservations about the reliability of the average achievement score because the percentage of students with achievement too low for estimation exceeds 15 percent, though it is less than 25 percent.
NOTE: Education systems are sorted by 2011 average percent correct. The answer shown illustrates the type of student response that was given full credit.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2011.

Page intentionally left blank

Appendix C: TIMSS-NAEP Comparison

How Does the Content of TIMSS 2011 Compare With That of the NAEP 2011 Mathematics and Science Assessments?

In reporting results on how U.S. students perform, the National Center for Education Statistics (NCES) draws on multiple sources of data in order to capitalize on the information presented in national and international assessments. In the United States, data on 4th-grade and 8th-grade students' mathematics and science achievement come primarily from two sources: the National Assessment of Educational Progress (NAEP) and the Trends in International Mathematics and Science Study (TIMSS). TIMSS provides internationally comparable data on student performance, while NAEP tracks performance nationally as well as in state and national subpopulations. This comparative study of TIMSS 2011 and NAEP 2009/2011 revealed important similarities and differences between the two assessments.

In the mathematics portion of the comparative study, the TIMSS 2011 and NAEP 2011 mathematics frameworks both specify five similar mathematical content areas to be assessed (*number, measurement, geometry, data*, and *algebra*). However, there are key differences between the two assessments. While both frameworks assess student performance in cognitive dimensions within these content areas, the dimensions are defined differently. For TIMSS, thethree dimensions assessed are *knowing*, *applying*, and *reasoning*, and each grade in TIMSS has a different distribution across these three cognitive domains. For NAEP, the three dimensions are *low*, *moderate*, and *high*, and the distribution of items across grade levels is fixed at 4th grade and 8th grade (25 percent, 50 percent, and 25 percent, respectively). In terms of item content, both TIMSS and NAEP emphasize number at 4th grade and shift the focus to algebra at 8th grade.

Item-by-item content match analyses of the mathematics assessments show a strong content correspondence between TIMSS 2011 and NAEP 2011. Only 1 percent of the 4th-grade items and 3 percent of the TIMSS 2011 8th-grade items could not be fit to a *specific* objective within the NAEP 2011 mathematics framework. Grade-level fit analyses revealed a slight mismatch between the two assessments. For TIMSS 4th-grade mathematics items, 11 percent were found to align better with the 8th grade in the NAEP mathematics framework. For TIMSS 8th-grade mathematics items, 14 percent were found to align better with the 4th grade or the 12th grade in the NAEP mathematics framework (3 percent and 11 percent, respectively), while 1 percent of the TIMSS 8th-grade items were found to be "no fit" at any grade level in the NAEP 2011 mathematics framework.[1] Finally, both TIMSS and NAEP assess students using multiple-choice and constructed-response item formats; however, TIMSS has a relatively equal proportion of items in the two formats at both grade levels, whereas NAEP has a greater proportion of multiple-choice items at both grade levels.

Turning to science, the TIMSS 2011 and NAEP 2009/2011 science frameworks assess similar content areas. However, there are key differences between the two assessments. For example, the two assessments differ in what science content is emphasized at each grade level. At 4th grade, when TIMSS items are mapped onto the NAEP science framework, TIMSS has the most items in *life science* (44 percent), followed by *physical science* (39 percent) and then *Earth and space sciences* (17 percent), compared with NAEP, which has a relatively equal distribution across the three content areas (each has about one-third of the items). At 8th grade, when TIMSS items are mapped onto the NAEP science framework, TIMSS has the most items in *physical science* (47 percent), followed by *life science* (34 percent) and then *Earth and space sciences* (18 percent), whereas NAEP places the greatest emphasis on *Earth and space sciences* (40 percent), followed by a nearly equal proportion of items in *physical science* (31 percent) and life science (30 percent).

Item-by-item content match analyses of the science assessments did not show the sort of strong content correspondence between TIMSS 2011 and NAEP 2011 found in the mathematics assessments. In science, 31 percent of TIMSS 4th-grade items and 23 percent of TIMSS 8th-grade items could not be fit with any content statement in the NAEP 2011 science framework. Grade-level fit analyses, likewise, revealed a degree of mismatch between the two assessments. For TIMSS 4th-grade science items, 22 percent were found to align best with the 8th grade in the NAEP science framework, while an additional 18 percent were considered "no fit" for either of the three grade levels (4th, 8th, or 12th) in the NAEP science framework. For TIMSS 8th-grade science items, 14 percent were found to align best with the 4th grade or 12th grade in the NAEP science framework (11 percent and 3 percent, respectively), while an additional 18 percent were considered "no fit" for either of the three grade levels (4th, 8th, or 12th)

1 The percentage of "no fit" items by grade level is smaller than the percentage of "no fit" items by content level. This is because some of the content-level "no fit" items were determined to be implied in the NAEP 2011 framework at a particular grade.

in the NAEP science framework.[2] Finally, both assessments use similar proportions of multiple-choice and constructed-response item formats, though the proportion of multiple-choice items is greater in NAEP than in TIMSS at both the 4th and 8thgrades.

In short, the purpose of the assessments, the content coverage, and the grade-level correspondence of the assessment items distinguish TIMSS 2011 from NAEP 2009/2011. The item differences are more noteworthy in the science assessments than in the mathematics assessments. Thus, it is important to bear in mind these differences when interpreting U.S. students' achievement, nationally and internationally, on NAEP and TIMSS.

2 As with the mathematics assessments, the percentage of "no fit" items by grade level in the science assessments is smaller than the percentage of "no fit" items by content level. This is because some of the content-level "no fit" items were determined to be implied in the NAEP 2011 framework at a particular grade.

Appendix D: Online Resources and Publications

Online Resources

The NCES website (http://nces.ed.gov/timss) provides background information on the TIMSS surveys, copies of NCES publications that relate to TIMSS, information for educators about ways to use TIMSS in the classroom, and data files. The international TIMSS website (http://www.timss.org) includes extensive information on the study, including the international reports and databases.

NCES Publications

The following publications are intended to serve as examples of some of the numerous reports that have been produced in relation to the Trends in International Mathematics and Science Study (TIMSS) by NCES. All of the publications listed here are available at http://nces.ed.gov/timss.

TIMSS 2007 Achievement Report

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., and Brenwald, S. (2008). *Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context* (NCES 2009-001 Revised). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001

TIMSS 2003 Achievement Report

Gonzales, P., Guzmán, J.C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., and Williams, T. (2004). *Highlights From the Trends in International Mathematics and Science Study (TIMSS) 2003* (NCES 2005-005). National Center for Education Statistics, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2005005

TIMSS 1999 Achievement Reports

Gonzales, P., Calsyn, C., Jocelyn, L., Mak, K., Kastberg, D., Arafeh, S., Williams, T., and Tsen, W. (2000). *Pursuing Excellence: Comparisons of International Eighth-Grade Mathematics and Science Achievement From a U.S. Perspective, 1995 and 1999* (NCES 2001-027). National Center for Education Statistics, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2001028

Gonzales, P., Calsyn, C., Jocelyn, L., Mak, D., Kastberg, D., Arafeh, S., Williams, T., and Tsen, W. (2000). *Highlights From TIMSS-R* (NCES 2001- 027). National Center for Education Statistics, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2001027

TIMSS 1995 Achievement Reports

National Center for Education Statistics, U.S. Department of Education. (1997). *Pursuing Excellence: A Study of U.S. Fourth-Grade Mathematics and Science Achievement in International Context (1995)* (NCES 97-255). National Center for Education Statistics, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=97255

Peak, L. (1996). *Pursuing Excellence: A Study of U.S. Eighth-Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context*: *Initial Findings From the Third International Mathematics and Science Study (1995)* (NCES 97-198). National Center for Education Statistics, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=97198

Takahira, S., Gonzales, P., Frase, M., and Salganik, L.H. (1998). *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context (1995)* (NCES 98-049). National Center for Education Statistics, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=98049

TIMSS Videotape Classroom Study Reports

Stigler, J.W., Gonzales, P., Kawanaka, T., Knoll, S., and Serrano, A. (1999). *The TIMSS Videotape Classroom Study: Methods and Findings From an Exploratory Research Project on Eighth-Grade Mathematics Instruction in Germany, Japan, and the United States (1995)* (NCES 1999-074). National Center for Education Statistics, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=1999074

National Center for Education Statistics, U.S. Department of Education. (2000). *Highlights From the TIMSS Videotape Classroom Study* (NCES 2000-094). National Center for Education Statistics, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2000094

Hiebert, J., Gallimore, R., Garnier, H., Givvin Bogard, K., Hollingsworth, H., Jacobs, J., Miu-Ying Chui, A., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., and Stigler, J. (2003). *Teaching Mathematics in Seven Countries: Results From the TIMSS 1999 Video Study* (NCES 2003-013 Revised). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2003013

Roth, K.J., Druker, S.L., Garnier, H., Lemmens, M., Chen, C., Kawanaka, T., Rasmussen, D., Trubacova, S., Warvi, D., Okamoto, Y., Gonzales, P., Stigler, J., and Gallimore, R. (2006). *Teaching Science in Five Countries: Results From the TIMSS 1999 Video Study* (NCES 2006-011). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid= 2006011

IEA Publications

The following publications are intended to serve as examples of some of the numerous reports that have been produced in relation to TIMSS by the IEA. All of the publications listed here are available at http://timss.bc.edu.

TIMSS 2011 Achievement Reports

Martin, M.O., Mullis, I.V.S., and Stanco, G.M. (2012). *TIMSS 2011 International Results in Science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. http://timssandpirls.bc.edu/isc/publications.html

Mullis, I.V.S., Martin, M.O., and Arora, A. (2012). *TIMSS 2011 International Results in Mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. http://timssandpirls.bc.edu/isc/publications.html

TIMSS 2007 Achievement Reports

Martin, M.O., Mullis, I.V.S., and Foy, P. (2008). *TIMSS 2007 International Science Report: Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/TIMSS2007/sciencereport.html

Mullis, I.V.S., Martin, M.O., and Foy, P. (2008). *TIMSS 2007 International Mathematics Report: Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/TIMSS2007/mathreport.html

TIMSS 2003 Achievement Reports

Martin, M.O., Mullis, I.V.S., González, E.J., and Chrostowski, S.J. (2004). *TIMSS 2003 International Science Report: Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss2003i/scienceD.html

Mullis, I.V.S., Martin, M.O., González, E.J., and Chrostowski, S.J. (2004). *TIMSS 2003 International Mathematics Report: Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss2003i/mathD.html

TIMSS 1999 Achievement Reports

Martin, M.O., Mullis, I.V.S., González, E.J., Gregory, K.D., Smith, T.A., Chrostowski, S.J., Garden, R.A., and O'Connor, K.M. (2000). *TIMSS 1999 International Science Report: Findings From IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss1999i/science_achievement_report.html

Mullis, I.V.S., Martin, M.O., González, E.J., Gregory, K.D., Garden, R.A., O'Connor, K.M., Chrostowski, S.J., and Smith, T.A. (2000). *TIMSS 1999 International Mathematics Report: Findings From IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss1999i/math_achievement_report.html

TIMSS 1995 Achievement Reports

Beaton, A.E., Martin, M.O., Mullis, I.V.S., González, E.J., Smith, T.A., and Kelly, D.L. (1996). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study.* Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss1995i/SciencB.html

Beaton, A.E., Mullis, I.V.S., Martin, M.O., González, E.J., Kelly, D.L., and Smith, T.A. (1996). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss1995i/MathB.html

Martin, M.O., Mullis, I.V.S., Beaton, A.E., González, E.J., Smith, T.A., and Kelly, D.L. (1997). *Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study.* Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss1995i/SciencA.html

Mullis, I.V.S., Martin, M.O., Beaton, A.E., González, E.J., Kelly, D.L., and Smith, T.A. (1997). *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss1995i/MathA.html

Mullis, I.V.S., Martin, M.O., Beaton, A.E., González, E.J., Kelly, D.L., and Smith, T.A. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss1995i/MathScienceC.html

TIMSS Technical Reports and Frameworks

Martin, M.O., Gregory, K.D., and Stemler, S.E. (2000). *TIMSS 1999 Technical Report*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss1999i/tech_report.html

Martin, M.O., and Kelly, D.L. (Eds.). (1996). *Third International Mathematics and Science Study Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss1995i/TechVol1.html

Martin, M.O., and Kelly, D.L. (Eds.). (1998). *Third International Mathematics and Science Study Technical Report, Volume II: Implementation and Analysis, Primary and Middle School Years*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss1995i/TechVol2.html

Martin, M.O., and Kelly, D.L. (Eds.). (1999). *Third International Mathematics and Science Study Technical Report, Volume III: Implementation and Analysis, Final Year of Secondary School*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss1995i/TechVol3.html

Martin, M.O., and Mullis, I.V.S. (Eds.). (2011). *TIMSS and PIRLS Methods and Procedures*. http://timssandpirls.bc.edu/methods/index.html.

Martin, M.O., Mullis, I.V.S., and Chrostowski, S.J. (2004). *TIMSS 2003 Technical Report*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss2003i/technicalD.html

Mullis, I.V.S, Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., Arora, A., and Erberber, E. (2005). *TIMSS 2007 Assessment Frameworks.* Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/TIMSS2007/frameworks.html

Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., González, E.J., Chrostowski, S.J., and O'Connor, K.M. (2003). *TIMSS Assessment Frameworks and Specifications 2003: 2nd Edition*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss2003i/frameworksD.html

Olson, J.F., Martin, M.O., and Mullis, I.V.S. (2008). *TIMSS 2007 Technical Report*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/TIMSS2007/techreport.html

TIMSS Encyclopedia

Mullis, I.V.S., Martin, M.O., Minnich, C.A., Stanco, G.M., Arora, A., Centurino, V.A.S., and Castle, C.E. (2012). *TIMSS 2011 Encyclopedia: Education Policy and Curriculum in Mathematics and Science, Volumes 1 and 2.* Chestnut Hill, MA: International Study Center, Boston College. http://timssandpirls.bc.edu/timss2011/encyclopedia-timss.html

Page intentionally left blank

Page intentionally left blank

www.ed.gov

ies.ed.gov