Research Brief

Synthesizing NAEP and International Large-Scale Assessment Score Trends: A Pattern of Diverging Performance

AIR®
Advancing Evidence.
Improving Lives.

Samantha Burg (NCES), Maria Stephens (AIR), and Lydia Malley (NCES)
October 2022

The Issue

Results from a single large-scale student assessment tend to focus on average scores in comparison to those from the previous administration. While important, such a focus can miss patterns over the long term and overlook results from other assessments. This brief integrates results from four national and international assessments to examine score trends over several time periods and percentiles of performance for 4th-graders, 8th-graders, and 15-year-olds in reading, mathematics, and science.

The Findings

Across the four assessments, there is a relatively consistent pattern: performance between high- and low-performing U.S. students is diverging. This pattern of divergence became prevalent over the last decade, occurring in almost every subject and grade combination, and it continues to be seen in almost all of the recent pre-pandemic comparisons.

The Implications

When the same pattern is seen over time and from multiple sources, the convergence of data points strengthens the argument that the pattern is valid. The pattern of score divergence described in this brief provides a baseline picture of U.S. student performance that is essential for understanding post-pandemic findings from studies that are soon to come.

About AIR

AIR is committed to increasing the effectiveness of education at all levels through rigorous research and evaluation, training, and technical assistance. Our work spans the learning lifespan—from pre-K to postsecondary education, career readiness, and adult education—and focuses on a wide range of topics, including STEM, social and emotional learning, and state and federal education policies.

Our rigorous, state-of-the-art research and evaluation work provides important evidence for education policymakers and practitioners to use when answering crucial questions about program implementation, challenges, and solutions.

www.air.org

The Issue

This brief looks across assessments to explore patterns in the score trends of U.S. students.

Nearly every year, the U.S. Department of Education releases data about U.S. student performance from the National Assessment of Educational Progress (NAEP) or one of the several international large-scale assessments (ILSAs) in which the United States participates. The individual assessment results are important and informative on their own, but given that they cover overlapping grades and subjects, the results can be integrated to extend the lessons learned from the study-specific reports and provide a more comprehensive look at U.S. student performance over time.

This Research Brief presents the results of a comprehensive study that compared score trends from NAEP, the Progress in International Reading Literacy Study (PIRLS), the Trends in International Mathematics and Science Study (TIMSS), and the Program for International Student Assessment (PISA). It includes results for U.S. 4th-graders, 8th-graders, and 15-year-olds in the core subjects of reading, mathematics, and science, focusing on trends in average scores as well as 90th- and 10th-percentile scores, which are used as thresholds for identifying high- and low-performing students.

For each assessment, the most recent available scores are compared to the scores from three previous administrations, thus identifying a long-term, intermediate, and recent trend for each. While exact time spans are assessment-specific, they generally correspond to about 20 years for long-term, 10 years for intermediate, and 2 to 5 years for recent trends. And because the most recent available scores for all assessments in the study are pre-pandemic (2019 and earlier), the results provide a valuable baseline for future comparisons.

For a comparative perspective, this brief also examines U.S. patterns in score trends against those of 8 other countries with similarly consistent participation in the ILSAs.

The research, data, figures, and findings in this summary publication are based on the AIR Research Note Synthesizing NAEP and International Large-Scale Assessment Score Trends: A Pattern of Diverging Performance.

Burg, S., Stephens, M., Malley, L., and Fonseca, F. (2022). *Synthesizing NAEP and International Large-Scale Assessment Score Trends: A Pattern of Diverging Performance.* Washington, DC: American Institutes for Research.

Go to AIR's web page on Large-Scale Assessment Score Trends.

Go to the NCES website for more information about NAEP and ILSAs.

Go to the AIR website for more information about our work with NAEP and ILSAs.

The Findings

Over the *long* term, there have been no declines in performance for U.S. students at the 10th percentile, average, or 90th percentile in any assessment in this study. U.S. students' most recent scores were the same as or higher than their scores from about 20 years earlier.

The most numerous and consistent increases in U.S. students' scores over the long term were in 4th- and 8th-grade mathematics, with NAEP showing increases across the distribution and TIMSS showing increases at the average and at the 90th percentile. U.S. students' scores in NAEP reading were also higher over the long term—at the average and at the 90th percentile for 4th-graders and at the 90th percentile for 8th-graders.

Changes in reading scores, however, were limited to NAEP, with the scores for U.S. 4th-graders in PIRLS and 15-year-olds in PISA showing no differences over the long term. There were no long-term differences in mathematics scores for 15-year-olds, per PISA, or in science scores, per TIMSS at both grades.

In sum, in 6 of 9 mathematics and reading assessments, high-performing students' scores were higher in the last administration before the pandemic than they were about 20 years earlier. Low-performing students' scores, in contrast, were higher in only two of these assessments, and score divergence between the top and bottom ends of the distribution was apparent in only four.

Assessment and subjects	Long-term trend ~20-year span		
	10th percentile	Average	90th percentile
Grade 4			
NAEP reading (2002–2019)	↔	↑	↑
PIRLS reading (2001–2016)	↔	↔	↔
NAEP mathematics (2003–2019)	↑	↑	↑
TIMSS mathematics (2003–2019)	↔	↑	↑
NAEP science[1]	—	—	—
TIMSS science (2003–2019)	↔	↔	↔
Grade 8/15-year-olds[2]			
NAEP reading (1998–2019)	↔	↔	↑
PISA reading (2000–2018)	↔	↔	↔
NAEP mathematics (2000–2019)	↑	↑	↑
TIMSS mathemtics (1999–2019)	↔	↑	↑
PISA mathematics (2003–2018)	↔	↔	↔
NAEP science[1]	—	—	—
TIMSS science (1999–2019)	↔	↔	↔
PISA science[3]	—	—	—

↑ Upward trend (Most recent score is higher than earlier year score $p < .05$)

↔ No change (Most recent score is not measurably different from earlier year score $p < .05$)

[1] The NAEP science framework was revised in 2009, so there is no long-term trend period to report in this study.

[2] NAEP and TIMSS results are for 8th-graders. PISA results are for 15-year-olds.

[3] The first year that science was administered as a "major" domain in PISA was 2006. However, since using that as a starting point for long-term trend would create a time span that is less than that for other assessments, we forgo a long-term data point in this study.

Over the *intermediate* term, a pattern of divergence between student performance at the 90th and 10th percentiles became prevalent.

In all assessments except NAEP 8th-grade science and PISA mathematics and science, the scores of U.S. high- and low-performing students diverged over the intermediate term. In some cases—PIRLS 4th-grade reading, PISA 15-year-old reading, and NAEP 4th-grade science—the scores at the 90th percentile rose over this time period while the scores at the 10th percentile did not change, indicating a top-led divergence not seen over the long-term period. In other cases—TIMSS 4th-grade mathematics and science—the scores at the 10th percentile dropped while the scores at the 90th percentile did not change, indicating a new bottom-led divergence.

However, in a number of assessments, divergence occurred because of changes at both ends of the distribution. In NAEP reading at both grades, NAEP mathematics at both grades, and TIMSS mathematics and science at grade 8, the 90th percentile score rose while the 10th percentile score dropped, indicating two-tailed divergence.

In sum, in 11 of 14 assessments, U.S. student performance diverged over the intermediate term, largely due to the emergent drops of low-performing students' scores and rises of high-performing students' scores in more assessments than shown over the long term.

Assessment and subjects	Intermediate trend ~10-year span		
	10th percentile	Average	90th percentile
Grade 4			
NAEP reading (2009–2019)	↓	↔	↑
PIRLS reading (2006–2016)	↔	↑	↑
NAEP mathematics (2009–2019)	↓	↔	↑
TIMSS mathematics (2011–2019)	↓	↔	↔
NAEP science (2009–2019)	↔	↑	↑
TIMSS science (2011–2019)	↓	↔	↔
Grade 8/15-year-olds[1]			
NAEP reading (2009–2019)	↓	↔	↑
PISA reading (2009–2018)	↔	↔	↑
NAEP mathematics (2009–2019)	↓	↔	↑
TIMSS mathematics (2011–2019)	↓	↔	↑
PISA mathematics (2009–2018)	↔	↔	↔
NAEP science (2009–2019)	↑	↑	↑
TIMSS science (2011–2019)	↓	↔	↑
PISA science (2009–2018)	↔	↔	↔

↑ Upward trend (Most recent score is higher than earlier year score $p < .05$)
↔ No change (Most recent score is not measurably different from earlier year score $p < .05$)
↓ Downward trend (Most recent score is lower than earlier year score $p < .05$)

[1] NAEP and TIMSS results are for 8th-graders. PISA results are for 15-year-olds.

Over the *recent* term, there were consistent declines in student performance at the 10th percentile, contributing to a persistent pattern of diverging scores.

In most assessments, the scores of U.S. high- and low-performing students also diverged over the recent term. At the 4th grade, this was entirely due to decreases in scores at the 10th percentile, which occurred in NAEP reading, TIMSS mathematics, and NAEP and TIMSS science. The scores of the 90th percentile 4th-graders, previously rising in most assessments, did not change over this time period.

Among older students, score decreases at the 10th percentile were also common—in NAEP 8th-grade reading, mathematics, and science, as well as in TIMSS mathematics and science. Except for NAEP reading (in which there were declines across the distribution), these decreases contributed to the divergent score pattern. In TIMSS, these decreases were coupled with increases at the 90th percentile, repeating the two-tailed divergence observed over the intermediate term. PISA was the only assessment of older students where the 10th percentile score did not decline over the recent term; however, U.S. students' scores still diverged in PISA reading and mathematics due to increases at the 90th percentile.

In sum, in 9 of 14 assessments across grades, the scores of low-performing U.S. students were lower in the last administration before the pandemic than they were about 2 to 5 years earlier. Of concern, in 5 assessments, this pattern of decline pulled down average scores as well.

Assessment and subjects	Recent trend ~2- to 5-year span		
	10th percentile	Average	90th percentile
Grade 4			
NAEP reading (2017–2019)	↓	↓	↔
PIRLS reading (2011–2016)	↔	↓	↔
NAEP mathematics (2017–2019)	↔	↑	↔
TIMSS mathematics (2015–2019)	↓	↔	↔
NAEP science (2015–2019)	↓	↓	↔
TIMSS science (2015–2019)	↓	↓	↔
Grade 8/15-year-olds[1]			
NAEP reading (2017–2019)	↓	↓	↓
PISA reading (2015–2018)	↔	↔	↑
NAEP mathematics (2017–2019)	↓	↓	↔
TIMSS mathematics (2015–2019)	↓	↔	↑
PISA mathematics (2015–2018)	↔	↔	↑
NAEP science (2015–2019)	↓	↔	↔
TIMSS science (2015–2019)	↓	↔	↑
PISA science (2015–2018)	↔	↔	↔

↑ Upward trend (Most recent score is higher than earlier year score $p < .05$)
↔ No change (Most recent score is not measurably different from earlier year score $p < .05$)
↓ Downward trend (Most recent score is lower than earlier year score $p < .05$)

[1] NAEP and TIMSS results are for 8th-graders. PISA results are for 15-year-olds.

International data show that the consistent pattern of diverging scores over the *intermediate* term was fairly unique to the United States.

In the United States, 4th-grade scores diverged in all three subjects over the intermediate term: bottom-led by decreases at the 10th percentile in mathematics and science and top-led by increases at the 90th percentile in reading. In contrast, among the 8 other education systems that participated in all years of the ILSAs included in this study, only two—Lithuania and New Zealand—had score divergence in more than one subject over the intermediate term at 4th grade. And only in New Zealand in reading was the divergence bottom-led. In fact, New Zealand was the only education system besides the United States where the scores of low-performing 4th-grade students declined in any subject over this time period.

Assessment and subjects	Top-led divergence 90th percentile ↑ 10th percentile ↔	Two-tailed divergence 90th percentile ↑ 10th percentile ↓	Bottom-led divergence 90th percentile ↔ 10th percentile ↓
Grade 4			
PIRLS reading	Hong Kong Hungary Lithuania Singapore **United States**		New Zealand
TIMSS mathematics	Lithuania New Zealand		**United States**
TIMSS science			**United States**
Grade 8/15-year-olds[1]			
PISA reading	Lithuania Russian Federation Singapore United Kingdom **United States**	Hong Kong	Hungary New Zealand
TIMSS mathematics	Hungary	**United States**	
PISA mathematics	England		
TIMSS science	Hungary	**United States**	England Hong Kong New Zealand
PISA science	Hungary Lithuania		

[1] TIMSS results are for 8th-graders. PISA results are for 15-year-olds.
NOTE: The intermediate term corresponds to 2006 to 2016 for PIRLS, 2011 to 2019 for TIMSS, and 2009 to 2018 for PISA. Italy was also included in this study but did not show score divergence in any of these grades or subjects. The United Kingdom participates in PISA, whereas the subnational entity of England participates in TIMSS; thus, the two are listed separately but considered together for the purposes of discussing cross-study patterns.

Among older students, U.S. students' scores diverged in all three subjects over the intermediate term: top-led in reading (15-year-olds) and two-tailed in mathematics and science (8th-graders). Across the other 8 education systems, only in Hungary and the United Kingdom/England was divergence observed across all three subjects, though sharing the same pattern as the United States only in reading. Hong Kong, Lithuania, and New Zealand had divergent scores in two subjects (reading and science) over this period.

Looking across education systems, the commonality is that score divergence in reading in the international assessments tended to be due to increases among high-performing students over the intermediate term. (This was true in over half of the education systems, including the United States.) However, the United States' specific patterns of score divergence in mathematics and science—and the prevalence of score divergence generally—appear to be unique among international peers.

The Implications

Multiple assessments confirm that U.S. students' scores have been diverging over time.

When the same pattern is seen repeatedly over time and particularly from multiple, independent sources, the convergence of information strengthens the argument that the pattern is valid. NAEP and ILSA results offer a unique opportunity to triangulate an emerging pattern—a divergence between the top and bottom ends of the score distribution in the United States. This widening pattern became prevalent over the past decade (the intermediate trend), occurring in almost every subject and grade combination, and it continues to be seen in almost all of the recent trend comparisons prior to the COVID-19 pandemic. Moreover, the international data indicate that this pattern is fairly unique to the United States. As post-pandemic results become available, it will be important to update this research and determine whether this pattern of divergence has been exacerbated, held steady, or improved. This study provides an important baseline for tracking the achievement of high- and low-performing students so that any equity concerns can be understood and addressed.

AIR®

AIR® Headquarters
1400 Crystal Drive, 10th Floor, Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**