TIMSS 2027

Progress Report

Matthias von Davier

TIMSS & PIRLS International Study Center

IEA General Assembly

Warsaw, Poland

October 5, 2026

TIMSS 2023 and TIMSS 2023 Longitudinal Activities & Accomplishments since September 2025

- *TIMSS 2023 Longitudinal International Results*
- TIMSS Insights reports based on TIMSS 2023 and TIMSS 2023 Longitudinal data

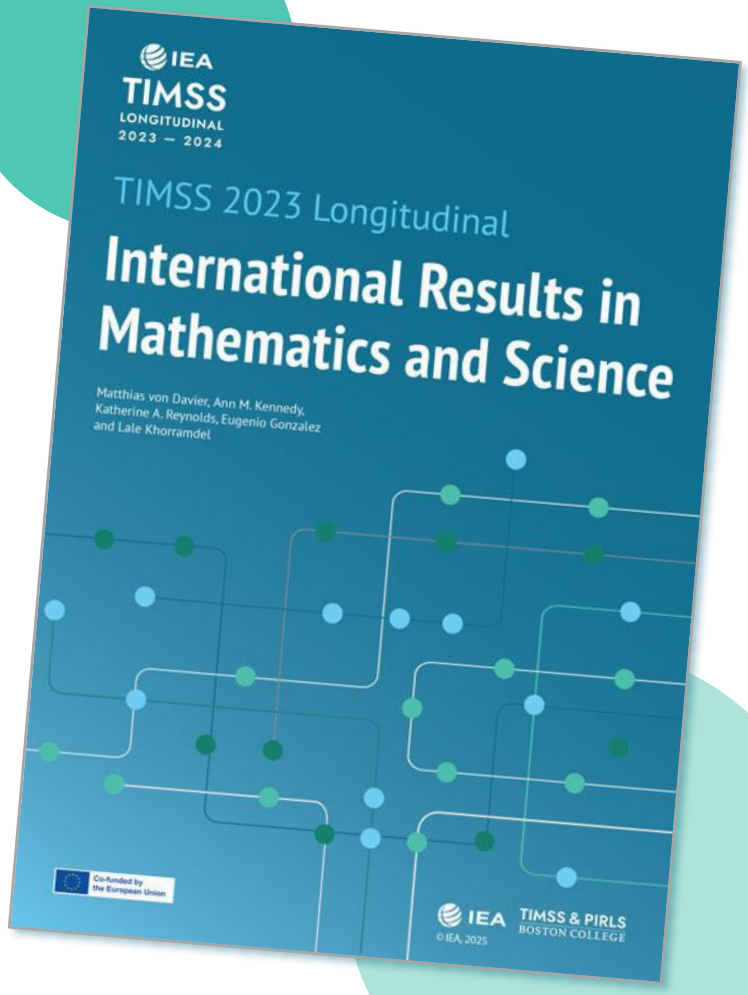
TIMSS Longitudinal International Results

Released **December 9, 2025**

- Live international release event in Podgorica, Montenegro

Highlights from the findings:

- More **educational resources at home** correlated with greater mathematics growth between grades
- Achievement improved more for student reporting a greater **sense of well-being at school**
- Frequent **absences or bullying** correlated with less growth and lower average achievement
- Fewer students had **positive attitudes** toward mathematics and science in Grade 5 than in Grade 4



TIMSS 2023 Insights Reports Since Last GA

https://timss2023.org/insights/

School Climate Health: Multidimensional Profiles of School Environments in TIMSS 2023

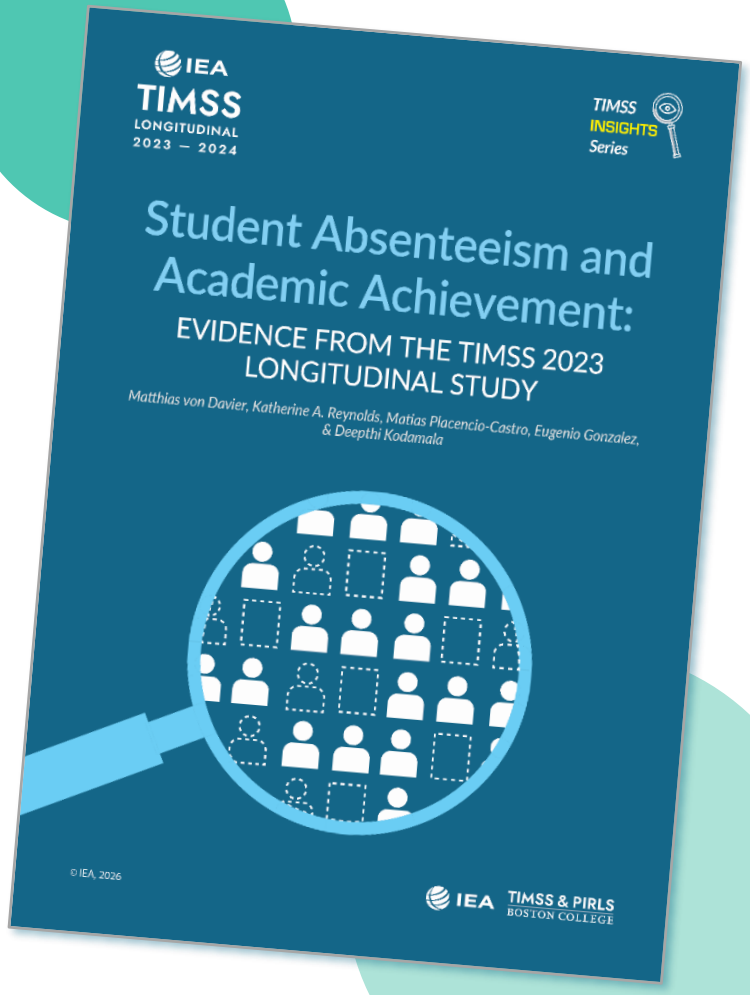
NOVEMBER 2025

Intended, Implemented, and Attained Mathematics and Science Curricula at the Eighth Grade: Findings from TIMSS 2023

MAY 2026

Student Absenteeism and Academic Achievement Growth: Evidence from the TIMSS 2023 Longitudinal Study

AUGUST 2026

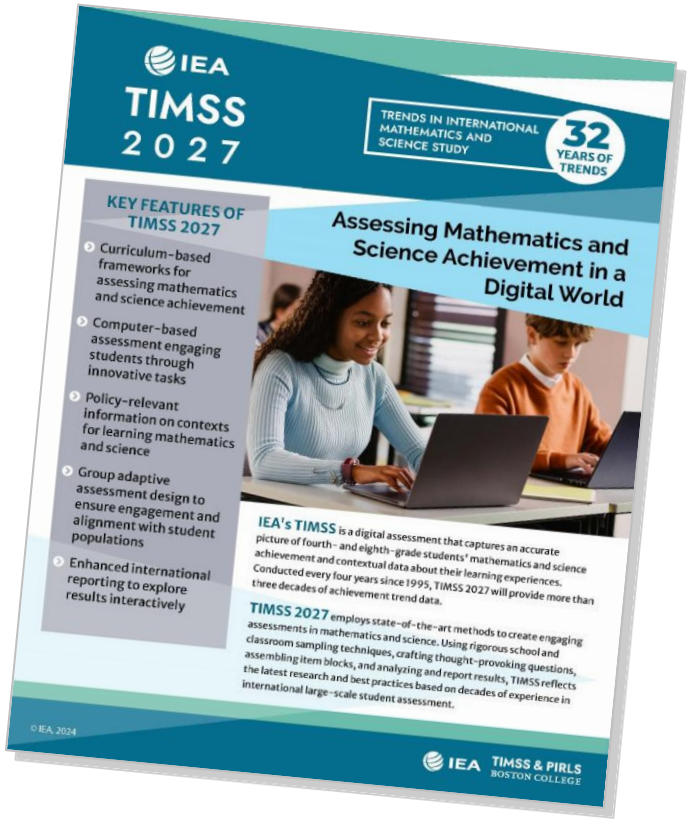


TIMSS 2027 Insights reports will follow the December 2028 release of the international results

TIMSS 2027

IEA's Trends in International Mathematics and Science Study

- The ninth cycle of TIMSS
- Fully digital assessment
- 65 countries and 6 benchmarking entities participating
- **Field test** in 2026
- **Main data collection** in 2027
- Report with international results in 2028



Grade 4
Grade 8

TIMSS 2027 Participants

- Albania
- Armenia
- Australia
- Azerbaijan
- Bahrain
- Belgium (Flemish)
- Belgium (French)
- Botswana (9)
- Brazil
- Bulgaria
- Canada
- Chile
- Chinese Taipei
- Côte d'Ivoire
- Croatia
- Cyprus
- Czech Republic
- Denmark
- England
- Finland
- France
- Georgia
- Germany
- Hong Kong SAR
- Hungary
- Iceland (5, 9)
- Iraq
- Ireland
- Israel
- Italy
- Japan
- Jordan
- Kazakhstan
- Korea, Rep. of
- Latvia
- Lithuania
- Macao SAR
- Malaysia
- Malta
- Montenegro
- Morocco
- The Netherlands
- New Zealand
- North Macedonia
- Northern Ireland
- Norway (5, 9)
- Oman
- Palestinian Nat'l Auth.
- Poland
- Portugal
- Qatar
- Romania
- Saudi Arabia
- Scotland
- Serbia
- Singapore
- Slovak Republic
- Slovenia
- Spain
- Sweden
- Türkiye (5)
- United Arab Emirates
- United States
- Uzbekistan
- Wales

Benchmarking

- Ontario, Canada
- Quebec, Canada
- South Africa (ICT; 6, 9)
- Abu Dhabi, UAE
- Dubai, UAE
- Sharjah, UAE

TIMSS 2027 Activities & Accomplishments

since September 2025

- Preparations for the Field Test
- Field Test administration
- Field Test data analysis and reviews
- Finalization of Data Collection instruments
- Preparations for automated scoring

Field Test Preparations

- **International versions** of instruments provided to countries in September 2025
- **Translation** with instructions provided in SOP manual; webinars on Assessment Master and StudyExpert
- Translation **verification** with cApStAn
- **Pilot study** for field-test scoring guide preparations October–December 2025
- Countries began receiving their **national players** in early 2026
- **International scoring training** (4th NRC meeting) in March 2026

Field Test March 1–April 30, 2026

MATHEMATICS

GRADE 4
128 items

GRADE 8
169 items

SCIENCE

GRADE 4
141 items

GRADE 8
182 items



GRADE 4
58 countries
1,685 schools
63,850 students

GRADE 8
40 countries
995 schools
45,245 students

CONTEXT QUESTIONNAIRES

64,000 parents

8,800 teachers

2,650 principals

64,000 students

Field Test Data Analysis & Review

Reviewed by the International Study Center in June

Achievement items evaluated for:

- Overall difficulty
- Discrimination
- Frequency of omitted and not-reached responses
- Consistency across scorers (for human-scored items)

Context questionnaire items evaluated for:

- Sufficient interrelatedness among scale items
- Adequate and comparable variability across countries
- Whether resulting scales related to achievement in expected ways

Expert Group Reviews

Reviewed by the QIRC and SMIRC in July

- Special attention to overall statistical quality
- Context Questionnaire reviews focused on new items and scales
- Achievement results evaluated on:
  - Overall block difficulty and timing
  - Distribution of items across content and cognitive domains
  - Distribution of item types within and across blocks
- Meetings resulted in recommendation of instruments with suggested revisions

Proposed instruments reviewed at **5th NRC meeting** in August

Finalization of Data Collection Instruments

- Final edits to questionnaires and achievement items updated in IEA's StudyExpert and RM's Assessment Master
- Created documentation to identify changes from the Field Test
- Finalized block assignments and established rules for routing individual students to very easy blocks to minimize non-response
- Prepared translation notes for revised item content
- Began preparing Survey Operations Procedures Manuals for the Data Collection phase
- Began preparing software for data management

Automated Scoring in TIMSS 2027

Builds on implementation into previous cycles

Rule-based automated scoring

- Items with **straightforward** scoring (multiple choice, numeric entry)
- Items with more **complex** scoring (interactive charts, some graphing)

Large language model (LLM) scoring

- International Study Center presented on LLM scoring of **written-response** items at 4th NRC Meeting
- Machine scoring achieved **high levels of agreement** with human scoring
- Builds on artificial neural network **quality control** used in TIMSS 2023
- Will be used in TIMSS 2027 **alongside** human and rule-based scoring

Automated Scoring (AS) Update

- Planned AI-based approach for automated scoring presented at **4th NRC meeting**
- Initial exploration of trend items using TIMSS 2023 data
  - Focusing on 1-point single text area and checkbox/dropdown items with open-ended responses
- Items were excluded from operational AS consideration when human-scoring reliability was **below 85%** and exact agreement was **below 90%** (Jung et al., 2026)
- Full-data runs are now in progress for items that met the initial criteria
- Borderline or excluded items are being reviewed, combining AI-generated revisions with content experts' feedback



Key Automated Scoring (AS) Results

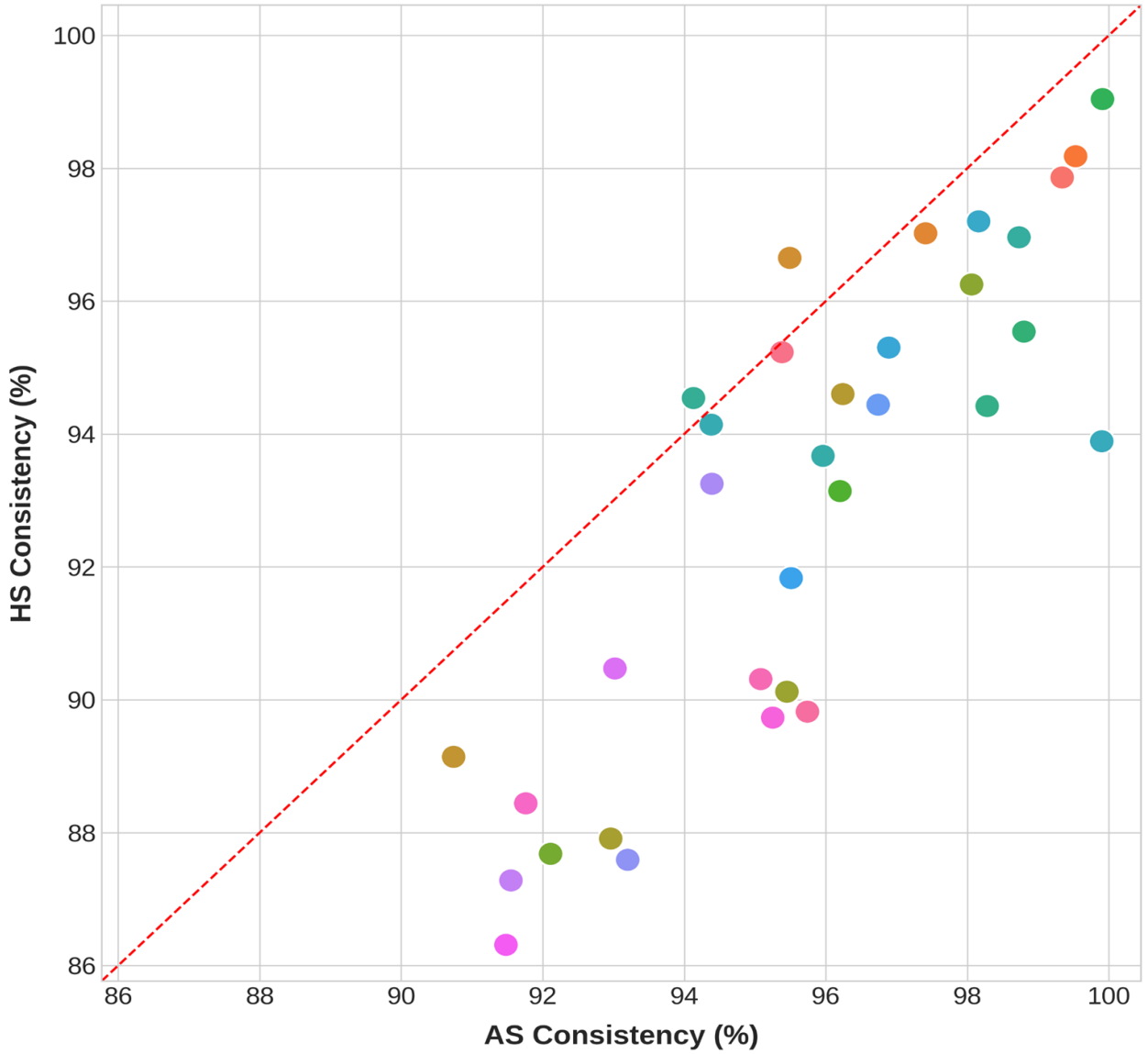
- Current results indicate strong alignment between automated and human scores across both subjects
- Overall AS reliability was **0.96**, exceeding the corresponding human-scoring reliability of **0.93**.

Subject	Num of Items	Agreement (%)	Cohen's Kappa	HS Reliability	AS Reliability
Math	8	95.96	0.90	0.95	0.96
Science	25	93.81	0.85	0.92	0.96
	33	**94.33**	**0.86**	**0.93**	**0.96**

Results reflect the current full-data runs; additional trend-item runs are still in progress.

Item-Level Reliability Checks

- For items with completed full-data runs, automated-scoring reliability is generally comparable to or higher than human-scoring reliability.
- Dots represent items, AS and HS agreement averaged across countries
- A similar picture emerges averaging across items and looking at countries.



Automated Scoring (AS) Summary

- Automated Scoring shows higher consistency/reliability than human scoring for the selected item set
- Human-Automated scoring agreement is as high as Human-Human scoring agreement
- We will be able to recommend a sizeable set of items for automated scoring
- Confirm the item set approved for operational AS
- Results will be communicated at **6th NRC Meeting** (Feb/Mar 2027) with presentation of the full AS pipeline

Upcoming Activities, Meetings, & Milestones in 2026

- **October 8:** Release of data collection instruments; countries make final necessary updates to translations/ national adaptations; SOP Unit 4: Sections on preparing context questionnaires and updating achievement test materials
- **October 22:** SOP Unit 3: Contacting Schools and Sampling Classes
- **October 29:** Within-school sampling software released
- **November 3–5:** Data Management Seminar (online)
- **November 12:** SOP Unit 5: Conducting the TIMSS 2027 Data Collection (including the Test Administrator Manual, National Quality Assurance Monitor Manual)



Upcoming Activities, Meetings, & Milestones in 2027

- **February 11:** Scoring the TIMSS 2027 Constructed Response Items
- **Feb 28–Mar 5:** International Scoring Training — 6th NRC Meeting
- **March - April:** 6th NRC Meeting Webinars: *TIMSS 2027 Encyclopedia*, Automated Scoring, Updates on Sampling, Operations, and Data Management
- **March 1–June 29:** **Northern Hemisphere data collection**
- **March 18:** SOP Unit 7: Creating and Submitting the TIMSS 2027 Databases and Documentation
- **March 25:** Final scoring guides and training materials
- **Aug 31–Dec 14:** **Southern Hemisphere data collection**
- **September:** Begin report planning for International Results



Thank you!

Matthias von Davier

TIMSS & PIRLS International Study Center

IEA General Assembly

Warsaw, Poland

October 5, 2026