28 Years of Trends

TIMSS 2023

Reaping the Benefits of a Fully Digital Assessment

Matthias von Davier, Ina V.S. Mullis, & Michael O. Martin

62nd IEA General Assembly - Online
October 2021

28 Years of Trends

TIMSS 2023: Innovations and Trend

- State of TIMSS 2023 Operations
- **Why are we doing all this?**
- Targeted Testing to Improve Measurement
- Leveraging Digital Assessment Delivery
- Process Data for Deeper Insights
- Quality Assurance to Support Measurement

TIMSS 2023: Innovations and Trend

28 Years of Trends

- State of TIMSS 2023 Operations
- **Why are we doing all this?**
- Targeted Testing to Improve Measurement
- Leveraging Digital Assessment Delivery
- Process Data for Deeper Insights
- Quality Assurance to Support Measurement

State of TIMSS 2023 Operations

- Project work advanced as planned despite COVID pandemic
- Country Participation
- Preparations for the Field test
  - Update of Assessment Frameworks
  - Update and Adaptation of Questionnaires
  - Development and Review of Items, Authoring + Implementation
  - Sampling Consultations and Preparations

28 Years of Trends

Countries Participating in TIMSS 2023

At the fourth grade:

- 61 countries
- 5 benchmarking participants

At the eighth grade:

- 46 countries
- 5 benchmarking participants

Item Development

Progress on TIMSS 2023 Assessment Development:

Frameworks were updated (multiple online meetings & reviews)

Items Developed for Field Test (with countries and partners)

- Fourth Grade - 22 blocks, 308 items
- Eighth Grade - 22 blocks, 330 items

Questionnaires were updated (multiple online meetings & reviews)

28 Years of Trends

TIMSS 2023: Innovations and Trend

- State of TIMSS 2023 Operations
- **Why are we doing all this?**
- Targeted Testing to Improve Measurement
- Leveraging Digital Assessment Delivery
- Process Data for Deeper Insights
- Quality Assurance to Support Measurement

TIMSS 2023: Innovations and Trend

28 Years of Trends

Why are we doing all this?

- TIMSS 2023 aims to be the gold standard of international trends assessment in design, delivery, and reporting
- TIMSS 2023 addresses the following challenges of international assessment using the advantages of computer-based testing:
  - a wide range of achievement distributions,
  - a need to cover science an mathematics topics broadly,
  - accounting for low stakes nature of the assessment
  - maintaining trend while adapting the assessment over time

TIMSS 2023: Innovations and Trend

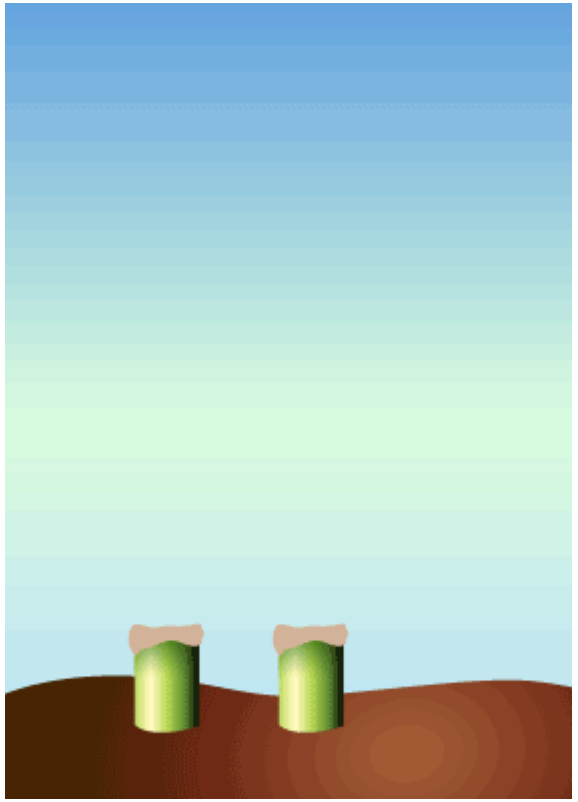
28 Years of Trends

- State of TIMSS 2023 Operations
- **Why are we doing all this?**
- Targeted Testing to Improve Measurement
- Leveraging Digital Assessment Delivery
- Process Data for Deeper Insights
- Quality Assurance to Support Measurement

TIMSS 2023 Targeted (Group-adaptive) Design

Goals of the targeted (group-adaptive) assessment:

- TIMSS 2023 as single, integrated assessment
- Better coverage of diverse range of achievement
- Increase student motivation, reducing frustration
- Reduce item-level nonresponse
- Minimize disruption of existing TIMSS design



TIMSS 2023 Targeted Design Essentials

The design maintains 14 item blocks for all countries

- Three levels of item block difficulty
  - Difficult (5) – Medium (4) – Easy (5)
- Item blocks combined into two levels of booklet difficulty
  - More difficult booklets (7)
    - Composed of difficult and medium item blocks
  - Less difficult booklets (7)
    - Composed of easy and medium item blocks

28 Years of Trends

Adapting the TIMSS 2023 Assessment

A **too easy test** will not measure more advanced achievement well

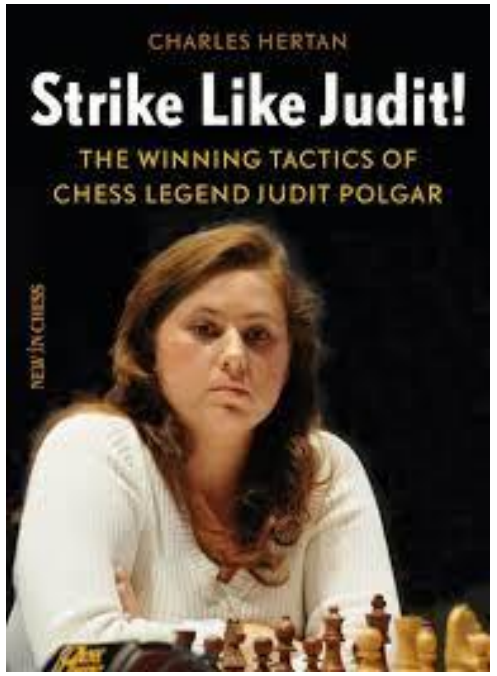
Still all are measured on the same **TIMSS** trend scale

Adapting the TIMSS 2023 Assessment

A **test of medium difficulty** measures average achievement well

Still all are measured on the same **TIMSS** trend scale

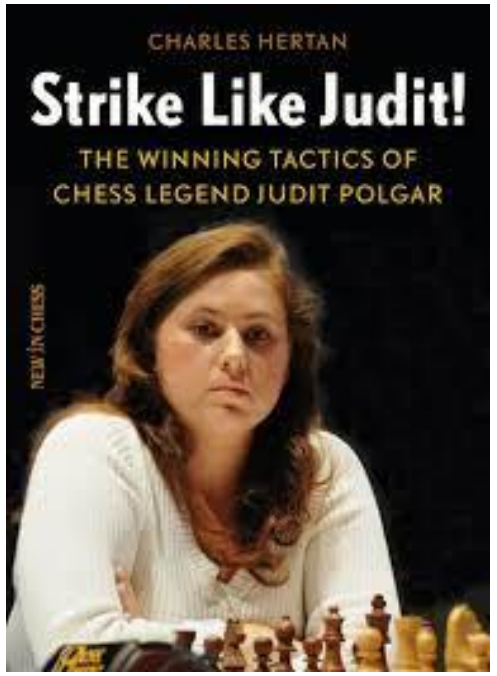
Adapting the TIMSS 2023 Assessment



A **too hard test** will not measure less advanced achievement well

Still all are measured on the same **TIMSS** trend scale

Adapting the TIMSS 2023 Assessment



A **targeted TIMSS 2023** measures **all** achievement levels well

Still all are measured on the same **TIMSS** trend scale

TIMSS 2023 Item Block Difficulty Targets

- **Long term** item block difficulty targets:
  - 40% for difficult item blocks
  - 55% for medium item blocks
  - 70% for easy item blocks
- The TIMSS 2023 trend item blocks are a mix of medium and difficult item blocks
  - TIMSS 2019 less difficult mathematics item blocks available as easy item blocks at grade 4
- The main target for TIMSS 2023 is developing easy item blocks
  - For 2023, **introduce 3 easy**, 1 medium, and 2 difficult item blocks to accelerate the development process

TIMSS 2023 Field Test

The field test administers new computer-based item blocks that help approaching the difficulty targets of the group-adaptive design

For each grade and subject:

- 5 easy item blocks to include 3 in the assessment
  - Use 5 less difficult item blocks for grade 4 mathematics
- 2 medium item blocks to include 1 in the assessment
- 4 difficult item blocks to include 2 in the assessment
  - A mix of PSIs and regular items

TIMSS 2023: Innovations and Trend

28 Years of Trends

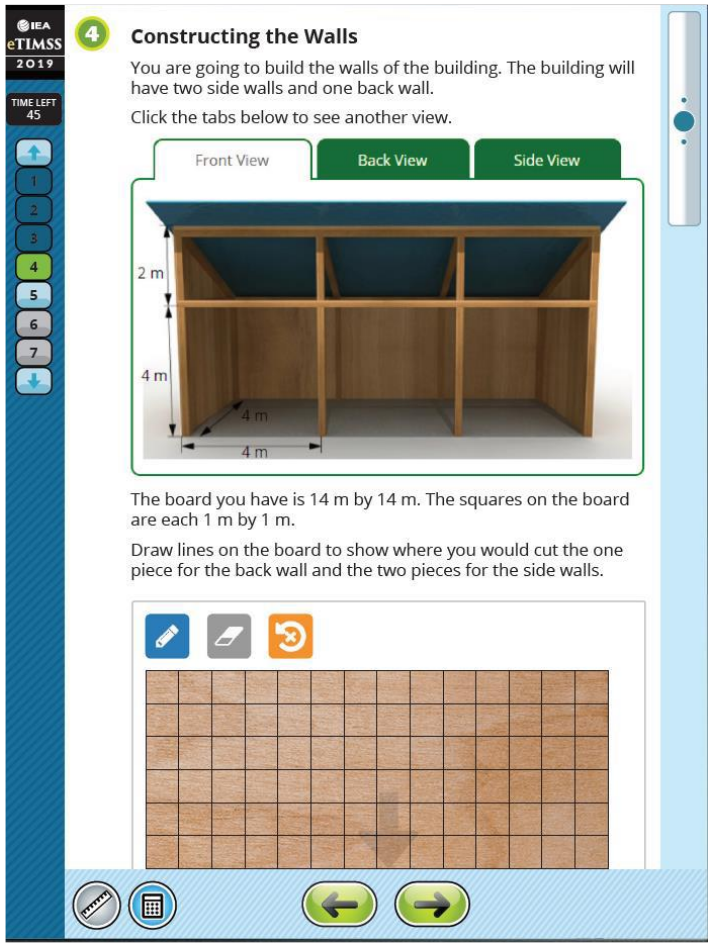
- State of TIMSS 2023 Operations
- **Why are we doing all this?**
- Targeted Testing to Improve Measurement
- Leveraging Digital Assessment Delivery
- Process Data for Deeper Insights
- Quality Assurance to Support Measurement

TIMSS 2023: Engaging & Interactive

A fully digital delivery of the TIMSS assessment takes item development to the next level:

- Computer based items can include animations, interactive elements and new response types
- This requires rethinking the limits imposed by paper based assessment
  - Multiple choice, constructed response distinctions are less obvious: drag and drop, draw, select, connect,…
  - Blocks, booklets, time limits and non-response can now be managed with assessment software tools

TIMSS 2023: Engaging & Interactive Problem Solving and Inquiry Tasks



- The 'building' is an example of engaging, multi-step, interactive tasks
- This TIMSS 2019 task was released and data will be made available for research
- TIMSS 2023 will contain the next generation of interactive tasks

TIMSS 2023: Engaging & Interactive

TIMSS 2023 provides full integration of traditional item types and innovative, interactive items:

- The distinction of Problem Solving and Inquiry (PSI) vs. traditional item is no longer needed
  - PSI items are integrated in the 2023 design, long (full) and short (half) block versions will be used
  - Traditional items can evolve and be made more engaging using added benefits of features unavailable on paper
  - This reflects the evolution of user experience and expectations of functionality in online environments

28 Years of Trends

TIMSS 2023: Innovations and Trend

- State of TIMSS 2023 Operations
- Targeted Testing to Improve Measurement
- Advantages of Digital Assessment Delivery
- Process Data for Deeper Insights
- Quality Assurance to Support Measurement

TIMSS 2023: Process & Timing Data

- Computer-based assessment produces a host of additional data on how students work through the items
- One central aspect is the time to response
  - Do students rush through the assessment? (Students going too fast cannot read and answer questions carefully)
  - Do students take more time for harder items? (Students who persist on more difficult problems have a higher likelihood of success)
  - Do students run out of time? (Students who linger on the first few problems or get stuck may not finish the assessment)

Making Use of Process Data

- Incorporate reporting process data into international reports – timing and event data
- Use timing and event data to examine response patterns
  - Explore test-taking strategies
  - Approaches to problem solving and inquiry
    - Evidence of strategy; more or less successful strategies
  - Identify disengaged and 'stuck' students

Making Use of Process Data

- Computer-based platforms record student interactions with the TIMSS and PIRLS assessments for quality control
  - Making sure test sessions can be completed and are on time
  - Helps finding issues in Field Test
- Computers can record more complex answers
  - In interactive items (PSIs) student responses tend to be 'authentic' (draw, drag, drop, write,…) computer tasks with many response modes
  - These complex responses are just like 'processes', they are interactive, and sequential, and open ended

Preparing for 2023 using TIMSS 2019 Process Data

- A research program on TIMSS and PIRLS process data is underway (Ph.D. students and postdoctoral researchers):
  - Timing and item revisits (navigation data) were being examined and results reported for TIMSS 2019 PSI
  - Non-response by item type, and together with timing data was examined
  - Results will be used to improve scoring of new items in TIMSS 2023,
  - The TIMSS 2023 field test will provide important first insights

TIMSS 2023: Process & Timing Data

- The TIMSS 2019 PSI report provided valuable insights into how students manage their time when responding to computer based extended tasks
  - The insights will be used to further improve the design and analyses of PSI blocks and regular items
  - Data on how students use the allowed time provides insights into which items are particularly challenging, which ones elicit immediate responses, and what types of items may lead some students to skip to the next item without responding

TIMSS 2023: Innovations and Trend

28 Years of Trends

- State of TIMSS 2023 Operations
- **Why are we doing all this?**
- Targeted Testing to Improve Measurement
- Advantages of Digital Assessment Delivery
- Process Data for Deeper Insights
- Quality Assurance to Support Measurement

TIMSS 2023: Quality Assurance Measures

- Computer based assessment has a range of advantages over paper based data collection
- Computers are reliable, do the same thing in a predictable way:
  - Saving the response data,
  - data transmission & cleaning,
  - scoring the data,
  - basic flagging of outliers and unexpected results
- Data will be of higher quality

TIMSS 2023: Quality Assurance Measures

- Quality control measures in a computer-based world goes beyond monitoring data collection in a subset of schools
- Changes over time & comparisons with past results allow improvement of overall data quality
- Scoring consistency, outlier detection and identification of unusual patterns is facilitated and data be cleaned more efficiently
- **A fully digital TIMSS 2023 benefits all participating countries**

TIMSS 2023: Automated Scoring Supports Quality Assurance

TIMSS 2023 utilizes technology for automated scoring of open responses:

- Supporting Human Scoring of Open Responses
- Automated Scoring of Graphical Responses
- Automated Scoring of Written Responses

TIMSS 2023: Automated Scoring for Quality Assurance

Automated Scoring of **Open Responses**

- Many more TIMSS items can be scored automatically in 2023
- Pattern matching based on gold standard responses
- Proximity to one of the possible correct responses is calculated
- Empty response or off-topic responses automatically detected
- Saves a lot of work that took place in countries

TIMSS 2023: Automated Scoring for Quality Assurance

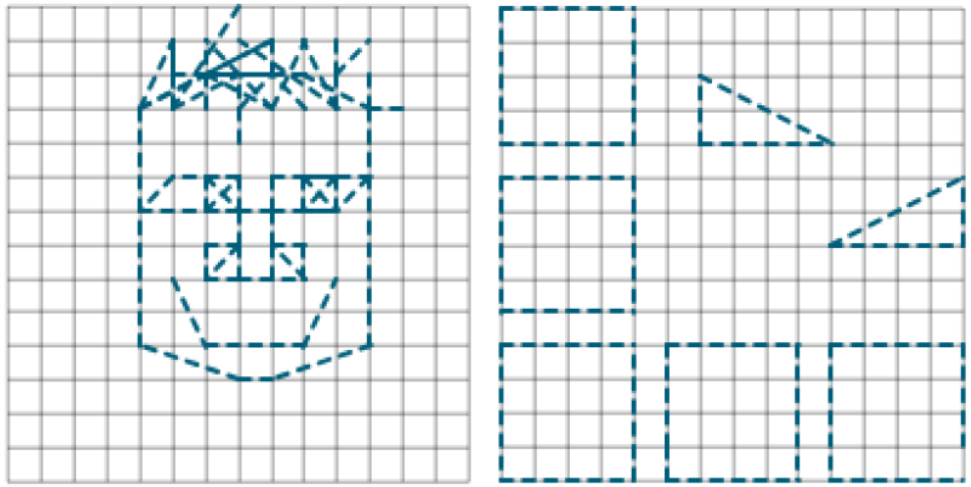
Supporting the Human Scoring of Open Responses

- Human scoring requires training of scorers
- Training can vary in success, scorers vary in quality
- Agreement between scorers can be better monitored
- Automated Scoring using AI can provide comparison data for scoring agreement
- Used in ETS e-Rater for scoring essays, in Duolingo English Test, and other testing programs

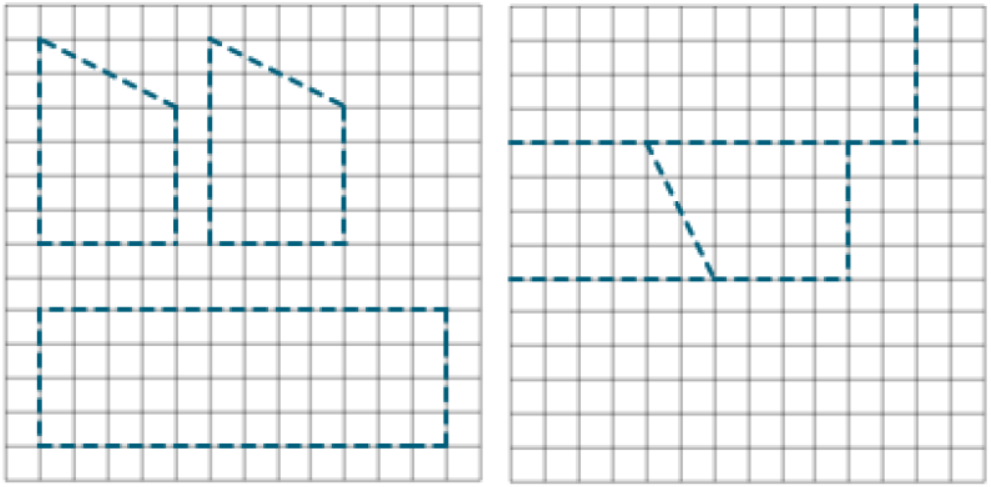
TIMSS 2023: AI Supported Scoring Research

Automated Scoring of **Graphical Responses**

Incorrect Responses



Correct Responses



TIMSS 2023: AI Supported Scoring Research

Automated Scoring of **Graphical Responses**

- Pilot project using artificial neural networks (ANNs)
- The project is training ANNs to score 15,000 line drawings collected in the 'Building' released PSI task
- Results are encouraging—very high agreement (between 91% and 98%) when comparing performance on test data (unseen by ANNs) with human scores
- IEA is supporting our work on a research project expanding the results to other TIMSS items

TIMSS 2023: AI Supported Scoring Research

28 Years of Trends

Automated Scoring of **Written Responses**

- Open ended responses from released items were analyzed in three languages: Korean, German, and English
- 'bag of words' based classification methods were used

```
##   [1] "Label"     "hide"     "eat"      "hold"     "tail"      "head"
##   [7] "shade"     "sun"      "heat"     "help"     "squirrel"  "surviv"
##  [13] "give"      "use"      "behavior" "bodi"     "can"       "cool"
##  [19] "act"       "like"     "cape"     "block"    "sunlight"  "make"
##  [25] "keep"      "will"     "light"    "feel"     "hot"       "much"
##  [31] "umbrella"  "shelter"  "go"       "eye"      "tri"       "get"
##  [37] "think"     "look"     "rock"     "day"      "behaviour" "shine"
##  [43] "way"       "find"     "food"     "cover"    "hit"       "predat"
##  [49] "gete"      "tale"     "big"      "fur"      "put"       "shad"
##  [55] "ground"    "protect"  "air"      "live"     "dri"       "squirel"
##  [61] "stay"      "environ"  "warm"     "let"      "desert"    "water"
##  [67] "dont"      "face"     "know"     "someth"   "nut"       "thing"
##  [73] "away"      "burn"     "die"      "thier"    "anim"      "mayb"
##  [79] "need"      "self"     "allow"    "provid"   "lift"      "enviro"
##  [85] "respond"   "chang"    "around"   "squrrel"  "sade"      "come"
##  [91] "might"     "cold"     "place"    "wont"     "squirrl"   "prevent"
##  [97] "want"      "ther"     "see"      "sinc"     "ray"       "less"
## [103] "also"      "realli"   "dessert"  "cooler"   "weather"   "dehydr"
## [109] "drink"     "touch"    "caus"
```

TIMSS 2023: AI Supported Scoring Research

Automated Scoring of **Written Responses**

- Pilot project using machine learning (ML) methods
- Comparing scoring agreement of different approaches, logistic regression, ANNs, random forest, etc.
- Results are encouraging, performance on test data (unseen by methods) compared with human scores, up to 91% agreement
- We examine multilingual aspects by studying differences between approaches using a variety of methods for language processing.

Summary: Automated Scoring Applications in TIMSS 2023

Automated Scoring of **Written and Graphical Responses**

- Many more items can be scored automatically in TIMSS 2023
- For Human Scored Extended Responses:
  - Machine scoring algorithms can be used for quality assurance
  - Agreement of automated scoring with human raters shows how well the items can be scored
  - Machine learning adds reassurance of scoring comparability

Presentation Summary: TIMSS 2023

Key innovations will be implemented in TIMSS 2023:

- A fully digital, targeted, more efficient and secure assessment will be delivered to countries
- Analysis and reporting, and researchers using the data collected in 2023 will benefit from these innovations
- Process data analysis and automated scoring using advanced technologies are two examples that will advance measurement

28 Years of Trends

TIMSS 2023

Reaping the Benefits of a Fully Digital Assessment

Matthias von Davier, Ina V.S. Mullis, & Michael O. Martin

62nd IEA General Assembly - Online
October 2021