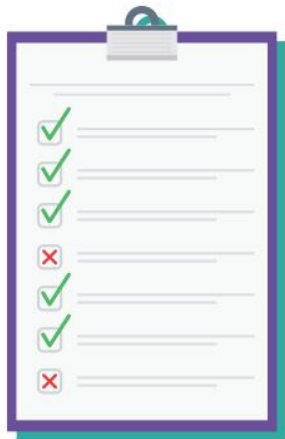


# TIMSS 2023: Progress Towards a Fully Digital Assessment



## Project Update

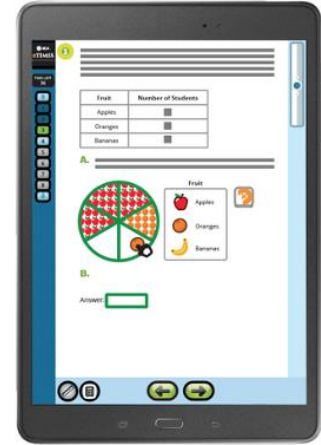
Matthias von Davier & Ina Mullis



General Assembly, Split,  
October, 2022

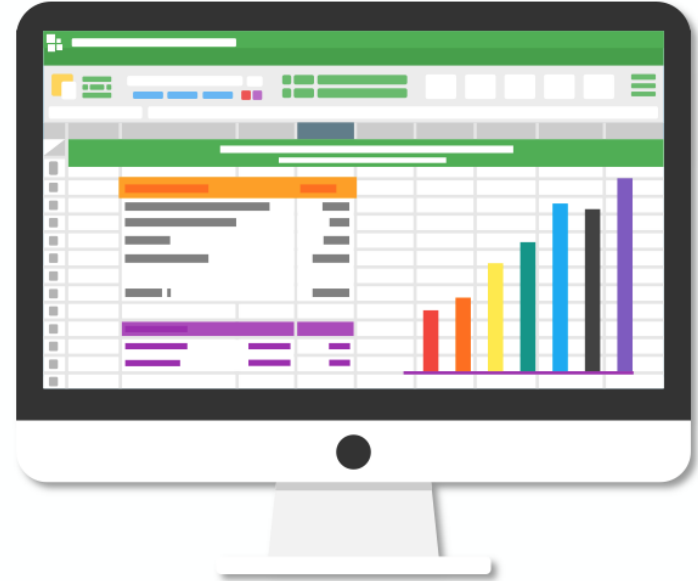
# TIMSS 2023 Project Update

- Field Test Data Collection and Analysis
  - Ensuring Proper Test Delivery
  - Getting the Data we Need
  - Analysing Field Test Data
  - Reviewing Results of the Field Test
- Preparing for the Main Data Collection
  - Finalizing Items
  - Assembly of the Main Data Collection Instruments
  - Ensuring Quality (Delivery System, Data Saving, Scoring)



# TIMSS 2023 Field Test Progress

- Field Test Activities
  - Ensuring Proper Test delivery
  - Getting the Data we Need
  - Analysing Field Test Data
  - Reviewing Items of the Field Test



# TIMSS 2023: Getting Ready for the Field

- Ensuring Proper Test Delivery
  - Digital Assessment needs a lot more centralized checks
  - Layout and Translation verification have a new sibling!
  - **Systems Verification! Checks whether:**
    - Software looks and works for students as planned
    - Saves and(!) exports all interactions properly

# System Verification – Scope of Material

## Achievement Items

### Field Test

- ❖ 512 items
  - 232 G4
  - 280 G8

### Data Collection

- ❖ 645 items
  - ~305 G4
  - ~340 G8

## PSIs

### Field Test

- ❖ 56 screens
  - 22 G4
  - 34 G8

### Data Collection

- ❖ 91 screens
  - ~41 G4
  - ~50 G8

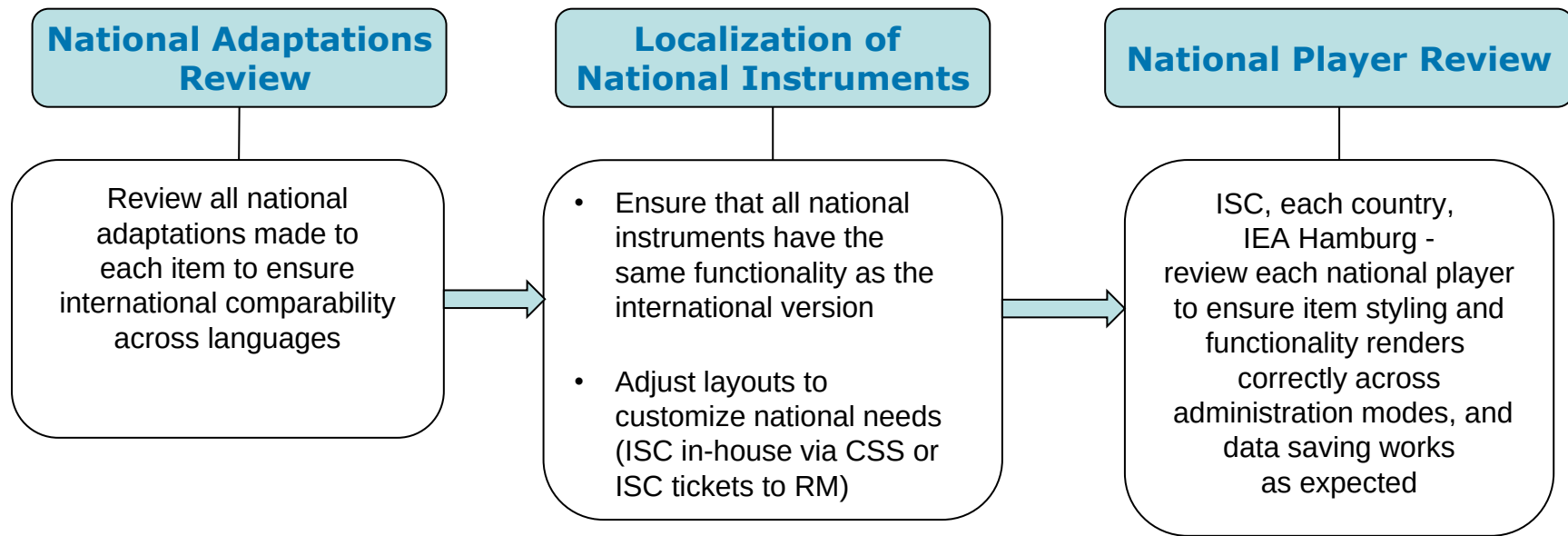
## Questionnaires

### 4 Questionnaires

- Home
- School
- Student
- Teacher

**Field Test:** 119 country languages  
**Data Collection:** 145 country languages

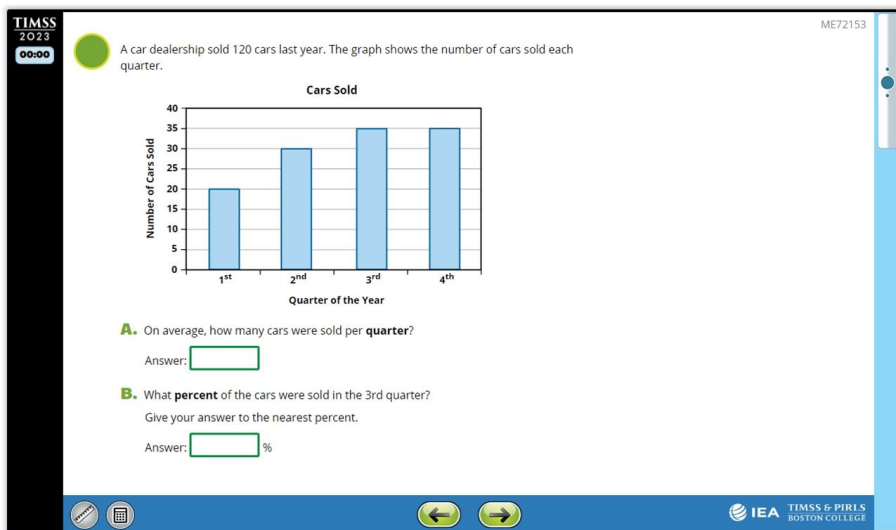
# Systems Verification – Activities



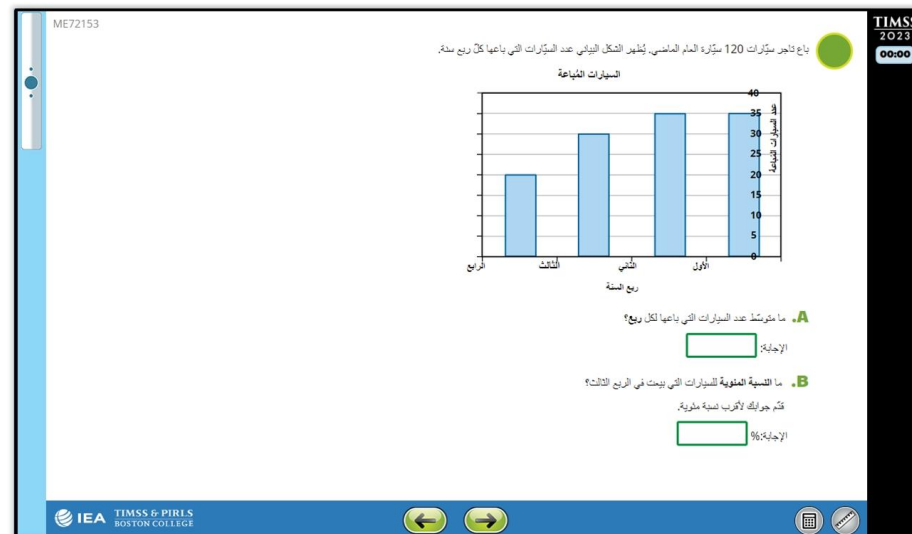
# Localization of National Achievement Items

## Example

### International version



### Arabic





# Localization of National PSIs

TIMSS  
2023  
00:00

M81E02 - Bees Screen 2



ハチ

ダンス中にまっすぐに飛んだ時間から、花のある場所までのきよりがわかります。

動画を見て、何秒まっすぐに飛んだか見てみましょう。

A. ハチは何秒まっすぐに飛んだでしょうか。

10の位までの<sup>くらい</sup>がい数に<sup>しやごにゆう</sup>四捨五入して答えてください。

答え:  秒

B. 花までのきよりは、1 秒に

この花は、何メートルのと

答え:  メー

Please adjust the alignment of MQ81B01\_T.Text1, the number pad box and MQ81B01\_T.Text2. This text should have the same alignment as part B below.



1 秒 = 800 メートル





# TIMSS 2023: Getting Ready for the Field

- Does the software save data as it should?
  - Opening, closing, using tools, entering special characters, drawing, erasing, dragging, highlighting,...
  - Students interact in many different ways, is all their input saved, is the last final answer saved properly, are timestamps correctly recorded,...
  - HUGE effort to check and recheck data saving, took longer than expected, but in the end we succeeded!

# TIMSS 2023: Getting the Data we Need

## Number of Achievement Items Field Tested

| Digital Field Test (New items) |             |         |       |
|--------------------------------|-------------|---------|-------|
| Grade                          | Mathematics | Science | Total |
| 4                              | 137         | 128     | 265   |
| 8                              | 169         | 177     | 346   |
| Total                          | 306         | 305     | 611   |

# TIMSS 2023: Getting the Data we Need

## Sample Sizes

| Grade        | Number of Countries | Number of Benchmarking Participants | Total |
|--------------|---------------------|-------------------------------------|-------|
| 4            | 48                  | 4                                   | 52    |
| 8            | 37                  | 2                                   | 39    |
| Total        | 85                  | 6                                   | 91    |
| Unique Total | 56                  | 4                                   | 60    |

# TIMSS 2023: Getting the Data we Need

## Sample Sizes

| Grade | Students | Parents | Teachers | Principals* |
|-------|----------|---------|----------|-------------|
| 4     | 55,638   | 51,811  | 3,761    | 1,587       |
| 8     | 45,162   |         | 6,786    | 1,082       |
| Total | 100,800  | 51,811  | 10,547   | 2,669       |

# Field Test Achievement Data Analysis

- The statistical and psychometric analyses of the field test data was carried out following best practices.
- Field test mathematics and science items were analysed across countries with respect to:
  - Comparability and level of challenge for adaptive design
  - Difficulty of correct responses and distribution of response options
  - Correlation of item responses and the targeted achievement scale
  - Scoring agreement across raters

# Field Test Context Data Analysis and Review

- The statistical and psychometric analyses of the context questionnaire data also followed best practices:
  - Appropriate response distribution of questions
  - Variability of context scales within and across countries
  - Internal consistency of context questionnaire scales
  - Correlation of context scales with achievement scales

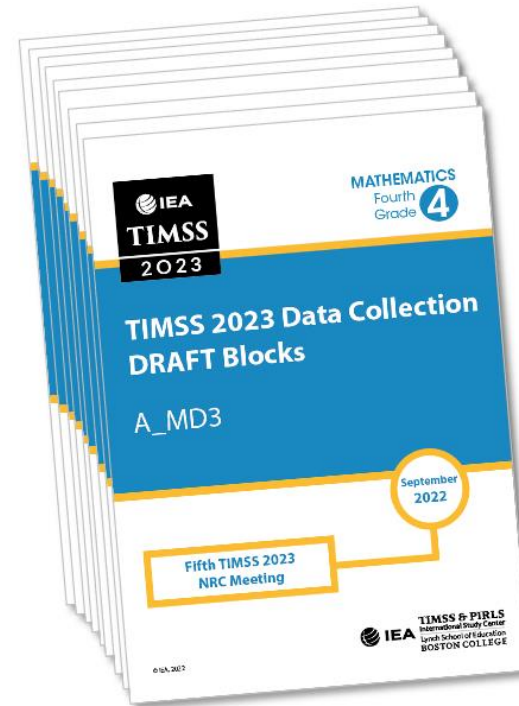
# Reviewing Items of the Field Test

- Items were reviewed guided by the statistical analysis results of the field test data
- Decisions based on item statistics (comparability across countries, overall psychometric quality) were confirmed during item review
- Some necessary changes were made based on item review observations (e.g. equation tool panel removal)
- Review showed a vast majority of items worked well and at the expected difficulty and discrimination levels



# Assembling the Main Data Collection Instruments

- Preparing for the Main Data Collection
- Group Adaptive Testing Design
- Trend Block Implementation
- New Block Assembly



# Group Adaptive Testing Design

**TIMSS 2023 group-adaptive design maintains 14 item blocks per grade and subject**

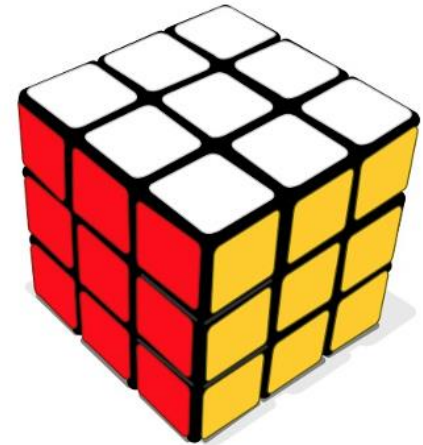
- Three levels of item block difficulty
  - Difficult (5) – Medium (4) – Easy (5)
- Item blocks combined into two levels of booklet difficulty
  - More difficult booklets (7)
    - Composed of difficult and medium item blocks
  - Less difficult booklets (7)
    - Composed of easy and medium item blocks



# TIMSS 2023 Design Essentials

## All countries administer all 14 booklets

- Higher performing countries:
  - Proportionally more of the more difficult booklets
- Lower performing countries:
  - Proportionally more of the less difficult booklets

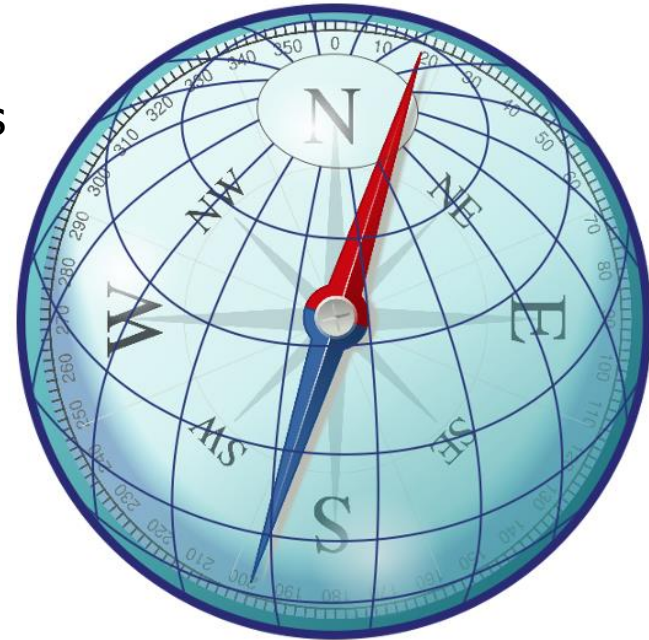


**This ensures all countries are compared using the same set of assessment blocks (covering the same content), but with different proportions for better targeting**

# Adapting the Assessment

## Our guiding principles:

- Most countries should assign equal proportions of the more and less difficult booklets
- Higher performing countries should assign more of the more difficult booklets (70%) and fewer of the less difficult booklets (30%)
- Lower performing countries should assign fewer of the more difficult booklets (30%) and more of the less difficult booklets (70%)



# Adapting the Assessment



# Trend Block Implementation

- Trend blocks from TIMSS 2019 are a subset of the TIMSS 2023 assessment to ensure sound measurement of achievement over time
- This means that trend blocks need to be carefully reimplemented as technology progresses,
  - to ensure that the same items given in TIMSS 2019 look and work in the same way
  - to allow countries that took TIMSS 2019 as a paper & pencil test to enjoy the benefits of computer based testing in TIMSS 2023
- The trend block assembly is completed, and countries are currently working on verifying and implementing the translations.



# Automated Test Assembly for new TIMSS blocks

- In any assessment program, assembling test forms based on a test blueprint or framework is a frequent challenge.
- Conventional test construction methods requires test developers to hand pick items from an item pool to create test forms satisfying all the requirements.
- This process depends on the test developer's experience and insights thus can be very subjective, time consuming and prone to error.
- Large-scale testing programs, such as TIMSS, utilize content and other constraints on imposed on multiple test forms to cover a wide range of content. This increases the difficulty of manually assembling tests.



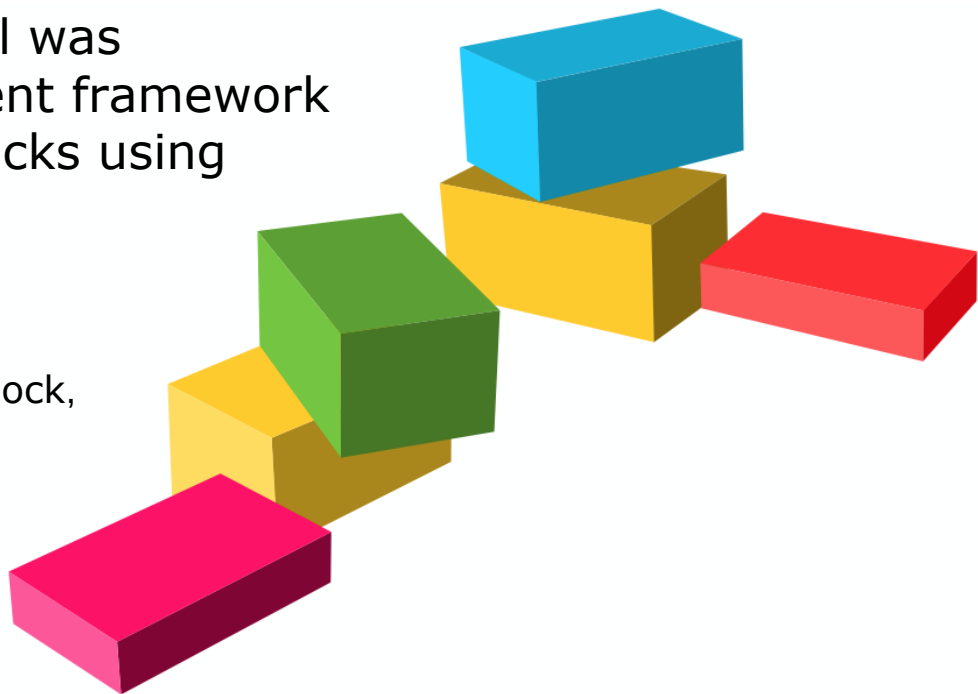
# Automated Test Assembly

- Automated test assembly (ATA) approach is proposed to increase efficiency of test construction with the help of computer algorithms.
- ATA approach use algorithms such as mixed-integer programming (MIP) to select a set of items from an item pool while adhering to all required constraints defined in the assessment framework.
- The process involves:
  - Build a model: Mathematical representation of defined constraints and objectives in a linear form.
  - Solve the model: Use mathematical programming solvers to find a globally optimal solution.



# TIMSS 2023 ATA

- For TIMSS 2023, the ATA model was constructed using the assessment framework specifications to form 6 new blocks using field test data.
- These include:
  - Number of blocks,
  - Minimum number of items per block,
  - Item inclusion constraint,
  - No item overlap constraint,
  - Cognitive domain constraints,
  - Content domain constraint,
  - Difficulty constraint,
  - Item type (MC, CR) constraint



## TIMSS 2023 ATA

- These constraints were usually specifications with fixed target values or upper and lower bounds. We also set a target deviation value for some of the constraints such as item type to avoid being overly restrictive.
- Separate models were defined for a specific grade (G4, G8) and subject (Mathematics, Science), and optimal solutions were obtained for each model.
- All analysis was carried out in R (R Core Team, 2020) and the ATA was implemented using eatATA package (Becker et al., 2021).

# TIMSS 2023 Expert Refined Test Assembly

- Subsequent to the draft blocks produced by the ATA algorithm, various refinements were implemented, and final blocks assembled
  - Items that were not selected but more engaging than selected items (e.g. animations and drag and drop vs. multiple choice) were used as replacement
  - Items that were providing context or potentially answers to other items on the same block were moved to other blocks
  - content coverage and processes, as well as difficulty and item discrimination levels were further improved where possible
  - item order was optimized, order of content domains emulates trend item blocks, order of difficulty within block was rearranged.

# Review of TIMSS 2023 Main Instruments

- All new blocks and questionnaires finalized and assembled for review
  - QIRC reviewed final updates to context questionnaires
  - SMIRC reviewed new blocks and commented
  - NRCs reviewed context questionnaires and new blocks
- Some additional change requests were made and are currently being implemented and booklets are being assembled
  - Combining trend and new blocks, layout verification
  - Functionality and data saving verification

# TIMSS 2023 Progress: Summary

- The TIMSS 2023 field test data collection and analysis was a success!
  - Despite technical challenges that can be expected in a rollout of this magnitude, the amount and quality of the field test data is outstanding, thanks to intensive QC and verification of all steps of the assessment platform production
- The analysis of the field test data showed that many of the newly developed items work well and can be used to assemble blocks for the country adaptive (targeted) test design in the main study
- The assembly of new high quality item blocks was facilitated by using human content and assessment expertise who finalized and refined blocks generated by algorithms for automated test assembly.

# Thank You!

Matthias von Davier & Ina Mullis

General Assembly, Split,  
October, 2022