unesco
Institute for Statistics

25 YEARS

Education Data and Statistics Commission

# Exemplar Foundational Mathematics Items

Leanne Ketterlin Geller, Ph.D.

Southern Methodist University, Dallas, TX

Global Alliance to Monitor Learning

#25YearsOfDataInsights

# Background: Need for Exemplar Foundational Mathematics Items

## Background

- To meet the criteria to report on SDG 4.1.1a, assessments must meet technical requirements, including specifications on the content and item difficulty
    - Calibrating item difficulty is important for accurate reporting
- The Technical Advisory Group (TAG) noted that additional guidance was needed to support meeting this criteria

## Purpose

- Examine item difficulty estimates from established assessments to understand item-level features that impact variations in difficulty
- Design example items that assess GPF Grade 2 across difficulty estimates for each subconstruct
- Please note the Technical Documentation has additional details

# Item-level Features Impacting Difficulty Estimates

## Construct Relevant Features

- Variations in the dimensions of the assessed content that the items are *intended* to measure
- Examples:
  - Number range
  - Familiarity
  - Complexity

## Construct Irrelevant Features

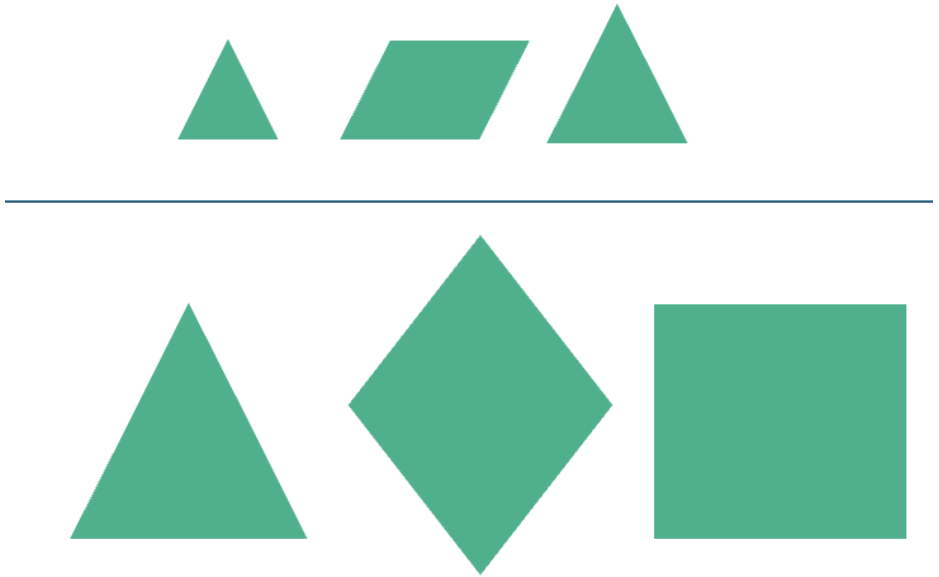
- Item-level features that the items is *not intended* to measure
- Examples:
  - Reading level of mathematics items
  - Visual representation
  - Item or response format

# G2.1: Geometry: Compose and decompose shapes and figures

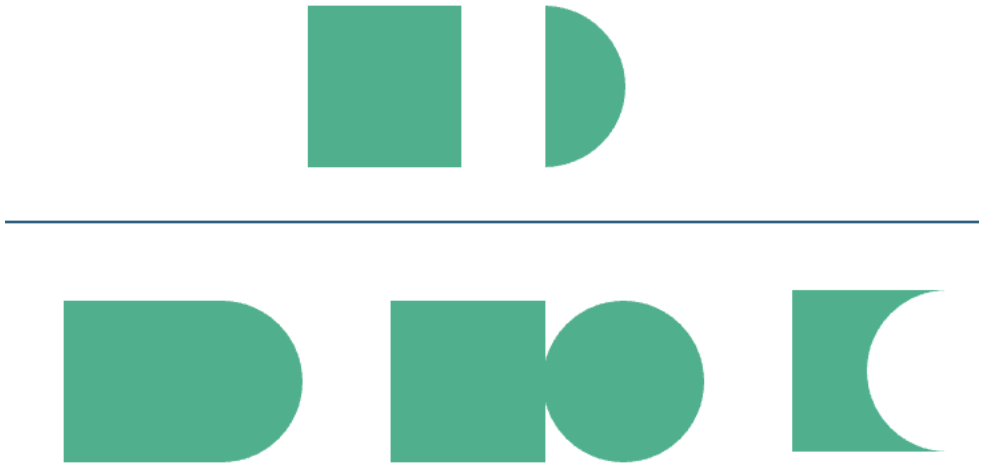
## Compose/decompose a larger two-dimensional shape from a small number of given shapes

**Which shape is made by putting these shapes together?**

**Difficult item**



**Less difficult item**



# Item Analysis from Existing Foundational Mathematics Assessments

Items and data obtained from administrations of:

- Early Grade Mathematics Assessment (EGMA)
- AMPL-ab
- UNICEF Foundational Learning Module (FLM) 2.0
- International Common Assessment of Numeracy (ICAN)

**Note: additional items and data are needed to improve the analyses

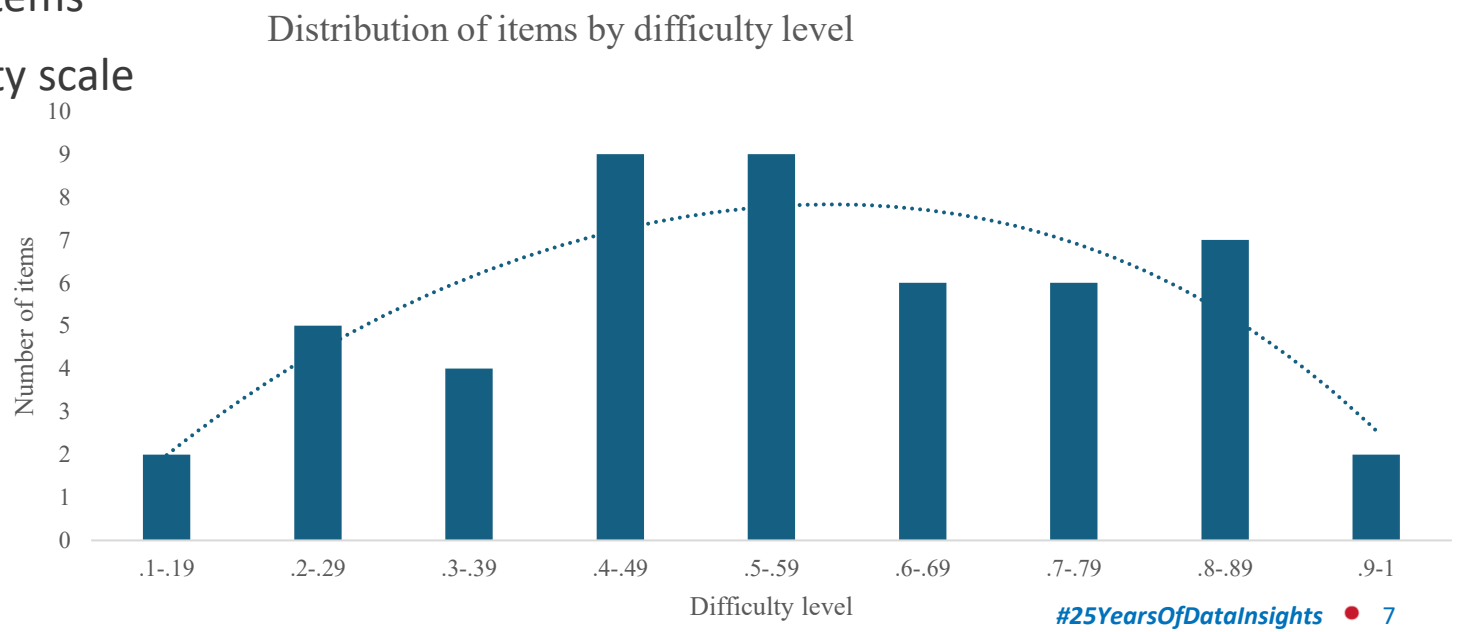
Analyses

- About 1,100 item statistics were examined
- *p*-value: proportion of correct responses to total attempted
  - Classical Test Theory estimation of item difficulty
  - Sample dependent
  - Low value: item is difficult for the sample
  - High value: item is less difficult for the sample
- Item-total correlation: indicator of alignment with construct
  - Items with values above 0.20 are included

| | Item Count by Grade | | | Item Difficulty Statistics (*p*-values) |
|---|---|---|---|---|
| Mathematics Subconstruct in Grade 2 of the GPF | 1 | 2 | 3 | Average *p*-value for G2-3 |

# Exemplar Items for Foundational Mathematics Constructs

- Findings were leveraged to design exemplar items
  - 50 items in total
- Test blueprint aligned with average item difficulty from existing foundational mathematics items
- Items represent ranges across the difficulty scale specified in the technical guidance
- At least 3 items per subconstruct



# Implications and Directions for the Future

## Future Test Development: New and Existing Tests

- Test Blueprint
  - Exemplar test blueprint may be used with the sample blueprints in the Technical Guidance to design the structure of a foundational mathematics assessment
- Exemplar items
  - Exemplar items may guide the development of foundational mathematics assessments to report on SDG 4.1.1a
  - Items assess all subconstructs at a range of difficulty levels

## Future Research and Limitations

- Additional items and data are needed to expand and enhance the set of exemplars
- Difficulty estimates were calculated using *p*-values
  - Sample dependent
  - Normative: influenced by the sample
    - Statistics may not generalize to all samples
    - Have limited value for making comparisons

unesco
Institute for Statistics

25 YEARS

Education Data and Statistics Commission

# THANK YOU

**Learn more:**

uis.unesco.org

databrowser.uis.unesco.org

@UNESCOstat

***#25YearsOfDataInsights***

Global Alliance to Monitor Learning