Automatic Item Generation: Experiences and Learning Points from Singapore Examinations and Assessment Board (SEAB)

Yick Chee FONG

at the 44th International Association for Educational Assessment (IAEA) Annual Conference, 9-14 September 2018, Oxford, United Kingdom

Abstract

In SEAB, we planned to develop a series of Computerised Adaptive Tests (CATs) on various Mathematics topics for Singapore elementary school students to assess their mathematical mastery at various points in their school journey. These CATs require large pools of items, which would be too costly and inefficient to produce using the traditional approach of engaging item-writers to write items one-by-one. Thus, SEAB adapted the automatic item generation approach to generate the required item pools, by first creating item templates from operationalised sources and then generating multiple items from each template using existing software such as SAS. With this approach, SEAB generated the items for two topical tests without incurring exorbitant costs. This paper shares SEAB's experience and learning points, as well as some findings on the item/template statistics, in adapting and implementing the automatic item generation approach.

Background

- **SEAB planned to develop a series of computerised adaptive tests (CATs) for formative assessment on Primary Mathematics topics**
- **CATs require large number of good items**
- **Writing items one-by-one is inefficient and not cost-effective**
- **Explored automatic item generation (AIG) to develop mathematics items**
  - Phase 1: item development for CAT Fractions
  - Phase 2: item development for CAT Decimals

Automatic Item Generation

A process of using item models to generate test items with the aid of computer technology.

Gierl, M., Lai, H., & Zhang, X. (2018). Automatic item generation. In *Encyclopedia of Information Science and Technology, Fourth Edition* (pp. 2369-2379). IGI Global.

SEAB's exploration on AIG focused on *cloning items* from existing resources (e.g. past exam items) by creating item templates

Item shells and item cloning

In item-cloning techniques (see, for instance, Bejar, 1993, or Roid & Haladyna, 1982) operational items are derived from "parent items" via one or more transformation rules. These parent items have been known as "item forms", "item templates", or "item shells", whereas the items generated from them are know now widely known as "item clones".

Scheerens, J., Glas, C. A., Thomas, S. M., & Thomas, S. (2003). *Educational evaluation, assessment, and monitoring: a systemic approach* (Vol. 13) (pp.108). Taylor & Francis.

Creating Item Templates (for CAT Fractions)

Phase 1 of our journey in exploring Automatic Item Generation

Source Item

Primary School Leaving Exam (PSLE) Maths item

Source: PSLE Maths

Level: Primary 5

Learning Objective: 2.6 solving word problems involving addition, subtraction and multiplication

A group of pupils took part in a quiz. $\frac{1}{3}$ of the boys and $\frac{1}{6}$ of the girls were prize-winners. There were 35 prize-winners altogether and $\frac{4}{7}$ of them were boys. What fraction of the pupils were prize-winners?

(1) $\frac{7}{27}$ (2)* $\frac{7}{30}$ (3) $\frac{7}{31}$ (4) $\frac{7}{34}$

Item Template (Stem & Options)

Item template (P5-T09)

A group of pupils took part in a quiz. $\frac{1}{a}$ of the boys and $\frac{1}{b}$ of the girls were prize-winners. There were *x* prize-winners altogether and $\frac{c}{d}$ of them were boys. What fraction of the pupils were prize-winners?

(1) $\frac{d}{ca+(d-c)b-3}$

(2)* $\frac{d}{ca+(d-c)b}$

(3) $\frac{d}{ca+(d-c)b+1}$

(4) $\frac{d}{ca+(d-c)b+4}$

Numbers are replaced by parameters (a, b, c, d) and formulae for the options.

In other word problems, parametrisation may also be done on words (e.g. names and objects in the problem context) which do not affect the mathematics of the problem.

Item Template (Constraints)

Item template (P5-T09)

Constraints on parameters

Possible values of a, b are: 2, 3, 4, 5, 6 (a, b take different values)

Possible values of c, d are: 2, 3, 4, 5, 6, 7, 8, 9 (c<d; c, d, a, b take different values)

x = 5d

(constraints ensure options are positive, proper fractions)

Constraints on options

Choose options such that all fractions are in simplest form

Arrange options in order

May adjust denominator beyond the values shown above (by integer increments)

Constraints are set based on curriculum and assessment considerations. Furthermore, constraints are set to rule out undesirable features (e.g. invalid or duplicate options).

Pros of stringent constraints: Better control of quality and difficulties of generated items

Cons of stringent constraints: Reduced number of items that can be generated

Generated items

Generated item (P5-T09-01)

a = 3, b = 6, c = 4, d = 7

A group of pupils took part in a quiz. $\frac{1}{3}$ of the boys and $\frac{1}{6}$ of the girls were prize-winners. There were 35 prize-winners altogether and $\frac{4}{7}$ of them were boys. What fraction of the pupils were prize-winners?

(1) $\frac{7}{27}$ (2)* $\frac{7}{30}$ (3) $\frac{7}{31}$ (4) $\frac{7}{34}$

Generated items

Generated item (P5-T09-02)
a = 2, b = 3, c = 4, d = 5

A group of pupils took part in a quiz. $\frac{1}{2}$ of the boys and $\frac{1}{3}$ of the girls were prize-winners. There were 25 prize-winners altogether and $\frac{4}{5}$ of them were boys. What fraction of the pupils were prize-winners?

(1)* $\frac{5}{11}$ (2) $\frac{5}{12}$ (3) $\frac{5}{13}$ (4) $\frac{5}{14}$

Generated items

Generated item (P5-T09-03)

a = 5, b = 4, c = 2, d = 3

A group of pupils took part in a quiz. $\frac{1}{5}$ of the boys and $\frac{1}{4}$ of the girls were prize-winners. There were 15 prize-winners altogether and $\frac{2}{3}$ of them were boys. What fraction of the pupils were prize-winners?

(1) $\frac{3}{10}$ (2) $\frac{3}{11}$ (3) $\frac{3}{13}$ (4)* $\frac{3}{14}$

Generated items

Generated item (P5-T09-04)
a = 5, b = 6, c = 8, d = 9

A group of pupils took part in a quiz. $\frac{1}{5}$ of the boys and $\frac{1}{6}$ of the girls were prize-winners. There were 45 prize-winners altogether and $\frac{8}{9}$ of them were boys. What fraction of the pupils were prize-winners?

(1) $\frac{9}{43}$ (2) $\frac{9}{44}$ (3)* $\frac{9}{46}$ (4) $\frac{9}{47}$

Adaptation of AIG Approach (for CAT Decimals)

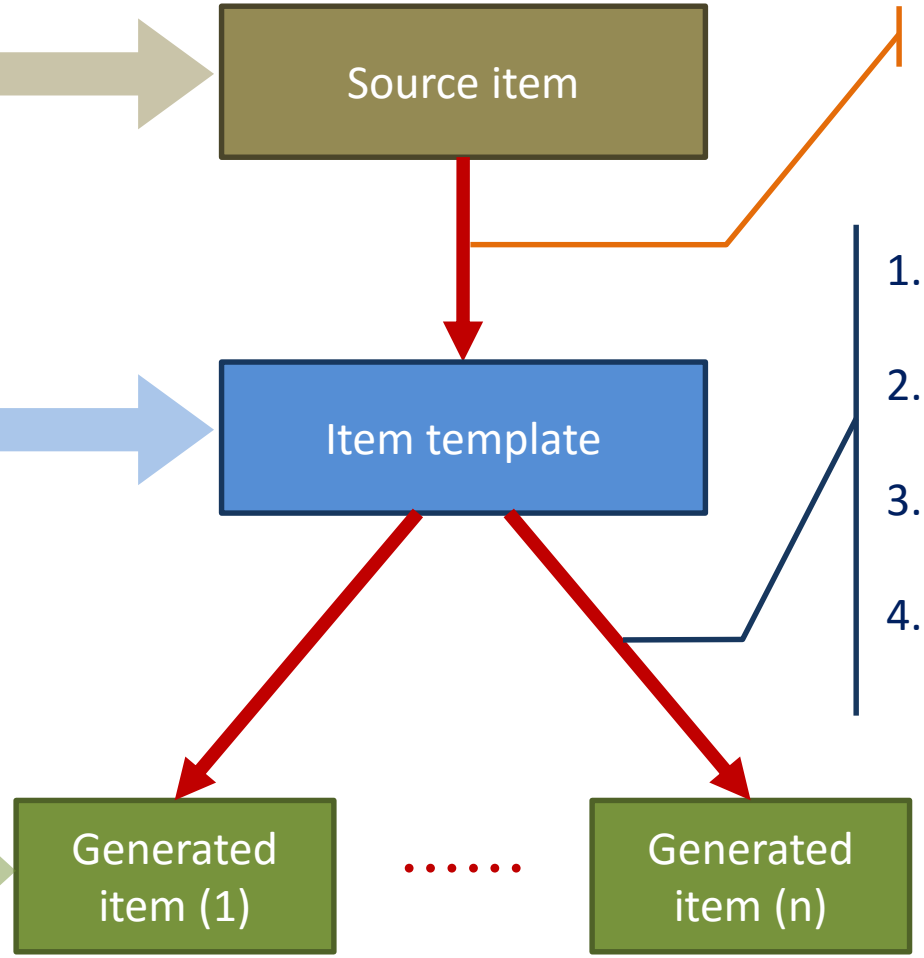
Phase 2 of our journey in exploring Automatic Item Generation

SEAB's adapted AIG approach

Existing resources such as maths items from past Primary School Leaving Examination (PSLE)

Item template specifies:
- Fixed and variable parts
- Rules for parameters
- Constraints for parameters

Pictorial form to be uploaded onto CAT platform



Convert source item to item template

(Implemented in Phase 1 and 2)

1. Write Statistical Analysis System (SAS) codes to generate item parameters
2. Export parameters, keys and IDs to comma-separated values (CSV) data
3. Create MS Word mail merge templates to read CSV data to display the items
4. Convert items from Word to PDF, and finally to JPEG for uploading

(Implemented in Phase 2)

Source Item

Primary School Leaving Exam (PSLE) Maths item

Source: PSLE Maths

Level: Primary 5

Learning Objective: 1.3 solving word problems involving the 4 operations

Jane cut a ribbon, 4.8 m long, into three pieces. The first piece was 0.9 m longer than the second piece. The second piece was 0.6 m shorter than the third piece. What was the length of the third piece of ribbon?

(1) 1.1 m

(2)* 1.7 m

(3) 2.1 m

(4) 3.3 m

Item Template (Stem & Options)

Item template (DECP5LO1.3_T004)

(Name) cut a ribbon, w m long, into three pieces. The first piece was x m longer than the second piece. The second piece was y m shorter than the third piece. What was the length of the third piece of ribbon?

(1) $((w - x - y) \div 3)$ m

(2)* $\{[(w - x - y) \div 3] + y\}$ m

(3) $((w + x + y) \div 3)$ m

(4) $(w - x - y) \div 3$ m

Item Template (Constraints)

Item template (DECP5LO1.3_T004)

Parameters

Name be female or male.

Female range: “Abby”, “Cailing”, “Fina”, “Nisha”, “Prema”, “Xiaowei”, “Yan” etc

Male range: “Adil”, “Ben”, “Caleb”, “Dinesh”, “Fadha”, “Glen” etc

Object range: “ribbon”, “string”, “wire”

Possible values of w, x, y = 0.1, 0.2, 0.3, …, 9.9

$w, x, y \neq 1, 2, 3, \ldots, 9; x \neq y; x, y < (w \div 3)$

10w and 10(x + y) are multiples of 3.

Constraints are then coded in SAS to generate the parameters for each item. Additional constraints may be specified to to rule out undesirable features (e.g. invalid or duplicate options).

Random Generation of Item Parameters (using SAS)

name item1 w x

(Name) cut a ribbon, w m long, into three pieces. The first piece was x m longer than the second piece. The second piece was y m shorter than the third piece. What was the length of the third piece of ribbon?

y

(1) ((w – x – y) ÷ 3) m pt1

(2)* {[(w – x – y) ÷ 3] + y} m pt2

(3) ((w + x + y) ÷ 3) m pt3

(4) (w – x – y) ÷ 3 m pt4

(constraints and SAS codes ensure options are positive numbers)

Parameters

Name be female or male.

Female range: “Abby”, “Cailing”, “Fina”, “Nisha”, “Prema”, “Xiaowei”, “Yan” etc

Male range: “Adil”, “Ben”, “Caleb”, “Dinesh”, “Fadha”, “Glen” etc

Object range: “ribbon”, “string”, “wire”

Possible values of w, x, y = 0.1, 0.2, 0.3, …, 9.9

w, x, y ≠ 1, 2, 3, …, 9; x ≠ y; x, y < (w ÷ 3)

10w and 10(x + y) are multiples of 3.

Randomly generate many sets of parameters

Obs	w	x	y	pt1	pt2	pt3	pt4	name	item1
1	3.3	0.4	0.8	0.7	1.5	1.5	2.1	Adil	string
2	4.8	0.1	0.8	1.3	2.1	1.9	3.9	Dinesh	ribbon
3	4.8	0.2	0.1	1.5	1.6	1.7	4.5	Fadha	wire
4	4.8	0.5	0.4	1.3	1.7	1.9	3.9	Nisha	ribbon
5	5.4	0.2	1.6	1.2	2.8	2.4	3.6	Fadha	wire
6	6.3	0.1	1.7	1.5	3.2	2.7	4.5	Fina	string
7	6.3	1.7	1.9	0.9	2.8	3.3	2.7	Cailing	wire
8	6.6	0.3	0.9	1.8	2.7	2.6	5.4	Abby	wire
9	6.6	1.1	0.4	1.7	2.1	2.7	5.1	Prema	wire
10	7.2	0.7	1.7	1.6	3.3	3.2	4.8	Glen	ribbon
11	7.8	0.3	1.2	2.1	3.3	3.1	6.3	Xiaowei	string
12	7.8	1.6	2.3	1.3	3.6	3.9	3.9	Abby	ribbon
13	7.8	1.8	0.9	1.7	2.6	3.5	5.1	Fina	ribbon
14	8.1	0.4	1.7	2.0	3.7	3.4	6.0	Adil	wire
15	8.1	1.6	0.2	2.1	2.3	3.3	6.3	Glen	wire
16	8.1	1.6	0.8	1.9	2.7	3.5	5.7	Caleb	ribbon
17	8.4	0.2	2.8	1.8	4.6	3.8	5.4	Cailing	wire
18	8.4	0.5	0.4	2.5	2.9	3.1	7.5	Glen	ribbon
19	8.4	1.9	0.8	1.9	2.7	3.7	5.7	Prema	ribbon
20	8.4	2.7	1.8	1.3	3.1	4.3	3.9	Cailing	wire

Export Item Parameters (from SAS to CSV)

- Select the desired number of "items" from the sets of randomly generated parameters (systematic sampling)
- Insert item ID, key and other useful variables
- Re-arrange the positions of key and options
- Export parameters as CSV data for MS Word mail merge

j	item_ID	key	w	x	y	opt1	opt2	opt3	opt4	name	item1	pron1	pron2	pron3
1	DECP5LO1.3_T004_201700001	2	4.2	1.1	0.4	0.9	1.3	1.9	2.7	Adil	ribbon	He	he	his
2	DECP5LO1.3_T004_201700002	2	7.2	0.2	1.9	1.7	3.6	3.1	5.1	Prema	wire	She	she	her
3	DECP5LO1.3_T004_201700003	3	7.8	2.3	0.4	5.1	3.5	2.1	1.7	Adil	ribbon	He	he	his
4	DECP5LO1.3_T004_201700004	3	8.4	1.7	0.1	6.6	3.4	2.3	2.2	Cailing	string	She	she	her
5	DECP5LO1.3_T004_201700005	2	9.9	2.1	3.3	1.5	4.8	5.1	4.5	Fadha	ribbon	He	he	his

Mail Merge Template (MS Word)

«j»

Item ID: «item_ID»
Key: («key»)

«name» cut a «item1», «w» m long, into three pieces. The first piece was «x» m longer than the second piece. The second piece was «y» m shorter than the third piece. What was the length of the third piece of «item1»?

(1) «opt1» m

(2) «opt2» m

(3) «opt3» m

(4) «opt4» m

Generated Items (MS Word → PDF → JPEG)

1

Item ID: DECP5LO1.3_T004_201700001
Key: (2)

Adil cut a ribbon, 4.2 m long, into three pieces. The first piece was 1.1 m longer than the second piece. The second piece was 0.4 m shorter than the third piece. What was the length of the third piece of ribbon?

(1) 0.9 m

(2) 1.3 m

(3) 1.9 m

(4) 2.7 m

Generated Items (MS Word → PDF → JPEG)

2

Item ID: DECP5LO1.3_T004_201700002
Key: (2)

Prema cut a wire, 7.2 m long, into three pieces. The first piece was 0.2 m longer than the second piece. The second piece was 1.9 m shorter than the third piece. What was the length of the third piece of wire?

(1) 1.7 m

(2) 3.6 m

(3) 3.1 m

(4) 5.1 m

Generated Items (MS Word → PDF → JPEG)

3

Item ID: DECP5LO1.3_T004_201700003
Key: (3)

Adil cut a ribbon, 7.8 m long, into three pieces. The first piece was 2.3 m longer than the second piece. The second piece was 0.4 m shorter than the third piece. What was the length of the third piece of ribbon?

(1) 5.1 m

(2) 3.5 m

(3) 2.1 m

(4) 1.7 m

Generated Items (MS Word → PDF → JPEG)

4

Item ID: DECP5LO1.3_T004_201700004
Key: (3)

Cailing cut a string, 8.4 m long, into three pieces. The first piece was 1.7 m longer than the second piece. The second piece was 0.1 m shorter than the third piece. What was the length of the third piece of string?

(1) 6.6 m

(2) 3.4 m

(3) 2.3 m

(4) 2.2 m

Generated Items (MS Word → PDF → JPEG)

5

Item ID: DECP5LO1.3_T004_201700005
Key: (2)

Fadha cut a ribbon, 9.9 m long, into three pieces. The first piece was 2.1 m longer than the second piece. The second piece was 3.3 m shorter than the third piece. What was the length of the third piece of ribbon?

(1) 1.5 m

(2) 4.8 m

(3) 5.1 m

(4) 4.5 m

Item Pools for CAT Fractions and CAT Decimals

Outputs from our journey in exploring Automatic Item Generation

Item Generation and Calibration

Item Generation	Phase 1 CAT Fractions	Phase 2 CAT Decimals
No. of item templates	56	100
No. of items generated	264	471
No. of learning objectives	24	15

Item Calibration (concurrent calibration using Rasch model)	Phase 1 CAT Fractions	Phase 2 CAT Decimals
No. of participants	≥ 2000 students (from P3 and P5)	≥ 4000 students (from P4 and P6)
Sample size per item	≥ 300	≥ 400

Learning Objectives assessed by CAT Fractions

Level	LO	LO Statement
P2	1.1	fraction as part of a whole
P2	1.2	notation and representations of fractions
P2	1.3	comparing and ordering fractions with denominators of given fractions not exceeding 12 • unit fractions • like fractions
P2	2.1	adding and subtracting like fractions within one whole with denominators of given fractions not exceeding 12
P3	1.1	equivalent fractions
P3	1.2	expressing a fraction in its simplest form
P3	1.3	comparing and ordering unlike fractions with denominators of given fractions not exceeding 12

Level	LO	LO Statement
P3	1.4	writing the equivalent fraction of a fraction given the denominator or the numerator
P3	2.1	adding and subtracting two related fractions within one whole with denominators of given fractions not exceeding 12
P4	1.1	mixed numbers, improper fractions and their relationships
P4	2.1	fraction as part of a set of objects
P4	3.1	adding and subtracting fractions with denominators of given fractions not exceeding 12 and not more than two different denominators
P4	3.2	solving up to 2-step word problems involving addition and subtraction

Curriculum Planning and Development Division. (2012). *Primary Mathematics Teaching and Learning Syllabus.* Ministry of Education, Singapore.

Learning Objectives assessed by CAT Fractions

Level	LO	LO Statement
P5	1.1	dividing a whole number by a whole number with quotient as a fraction
P5	1.2	converting fractions to decimals
P5	2.1	adding and subtracting mixed numbers
P5	2.2	multiplying a proper/improper fraction and a whole number **without calculator**
P5	2.3	multiplying a proper fraction and a proper/improper fractions **without calculator**
P5	2.4	multiplying two improper fractions
P5	2.5	multiplying a mixed number and a whole number
P5	2.6	solving word problems involving addition, subtraction and multiplication

Level	LO	LO Statement
P6	1.1	dividing a proper fraction by a whole number **without calculator**
P6	1.2	dividing a whole number/proper fraction by a proper fraction **without calculator**
P6	1.3	solving word problems involving the 4 operations

Curriculum Planning and Development Division. (2012). *Primary Mathematics Teaching and Learning Syllabus.* Ministry of Education, Singapore.

Learning Objectives assessed by CAT Decimals

Level	LO	LO Statement
P3	1.1	adding and subtracting money in decimal notation
P3	1.2	solving word problems involving addition and subtraction of money in decimal notation
P4	1.1	notation, representations and place values (tenths, hundredths, thousandths)
P4	1.2	comparing and ordering decimals
P4	1.3	dividing a whole number by a whole number with quotient as a decimal
P4	1.4	converting decimals to fractions
P4	1.5	converting fractions to decimals when the denominator is a factor of 10 or 100

Level	LO	LO Statement
P4	1.6	rounding off decimals to • the nearest whole number • 1 decimal place • 2 decimal places
P4	2.1	adding and subtracting decimals (up to 2 decimal places)
P4	3.1	multiplying and dividing decimals (up to 2 decimal places) by a 1-digit whole number
P4	3.2	solving up to 2-step word problems involving the 4 operations
P4	3.3	rounding off answers to a specified degree of accuracy

Curriculum Planning and Development Division. (2012). *Primary Mathematics Teaching and Learning Syllabus.* Ministry of Education, Singapore.

Learning Objectives assessed by CAT Decimals

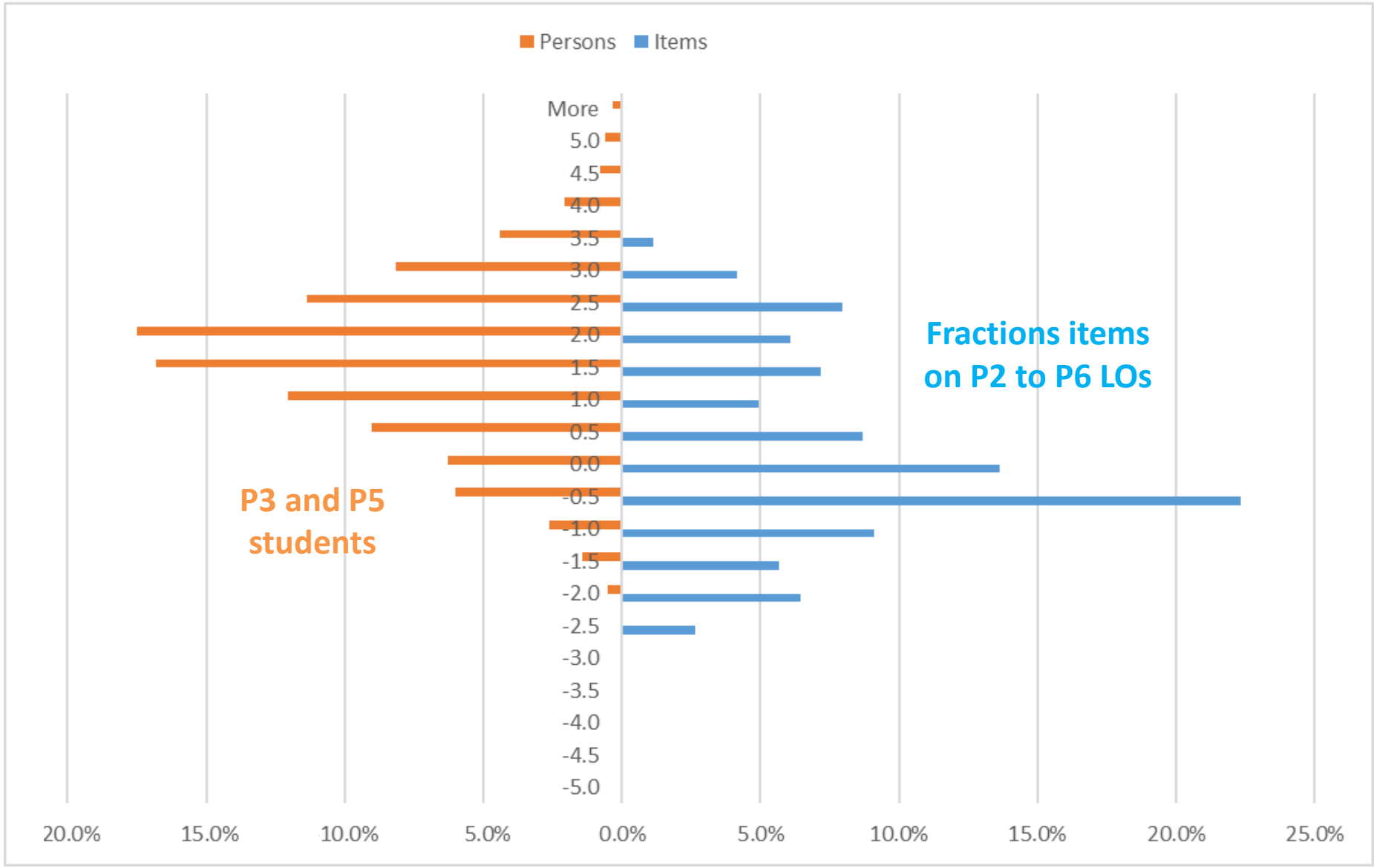
Level	LO	LO Statement
P5	1.1	multiplying and dividing decimals (up to 3 decimal places) by 10, 100, 1000 and their multiples **without calculator**
P5	1.2	converting a measurement from a smaller unit to a larger unit in decimal form, and vice versa • kilometres and metres • metres and centimetres • kilograms and grams • litres and millilitres
P5	1.3	solving word problems involving the 4 operations

Curriculum Planning and Development Division. (2012). *Primary Mathematics Teaching and Learning Syllabus.* Ministry of Education, Singapore.

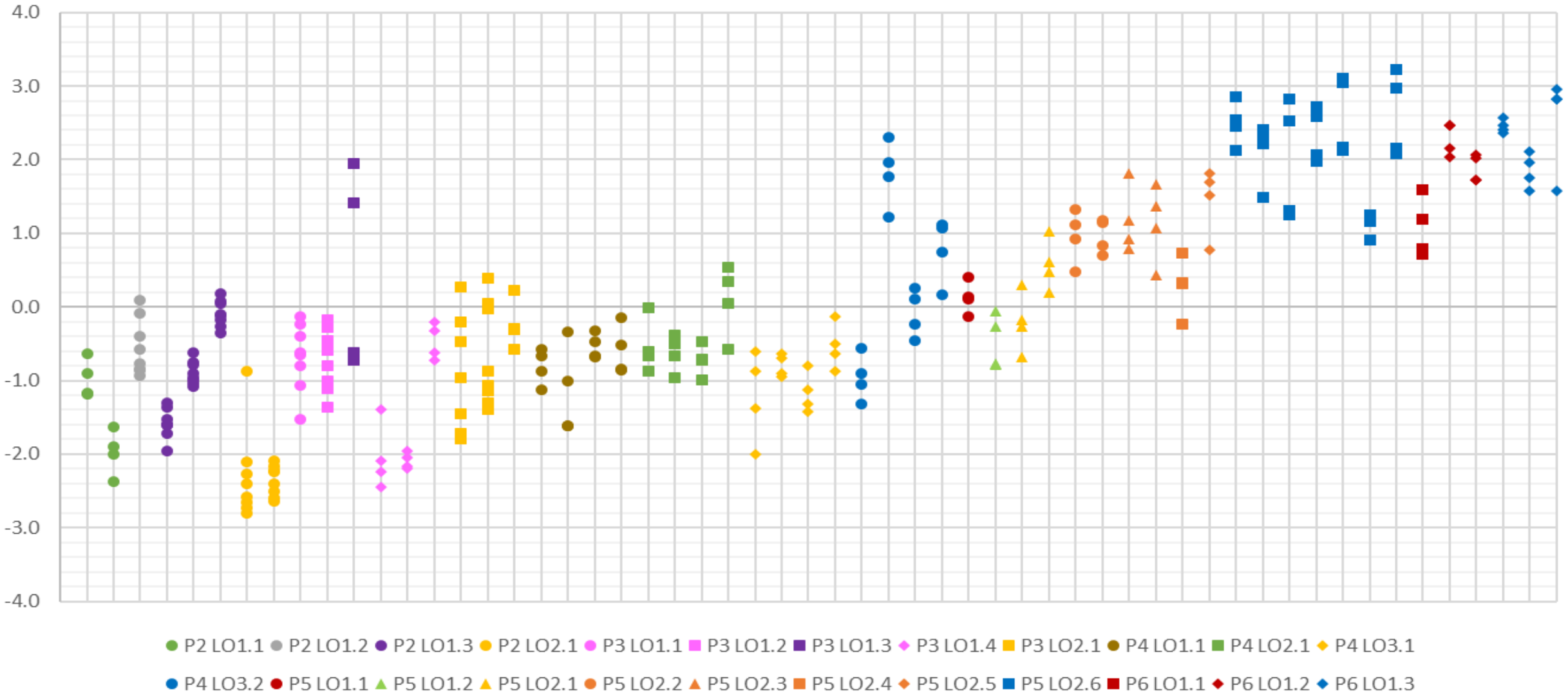
Empirical Results

How good are the items and templates?
What are some learning points from the use of item templates?

CAT Fractions: Person vs Item Distributions (Wright Map)

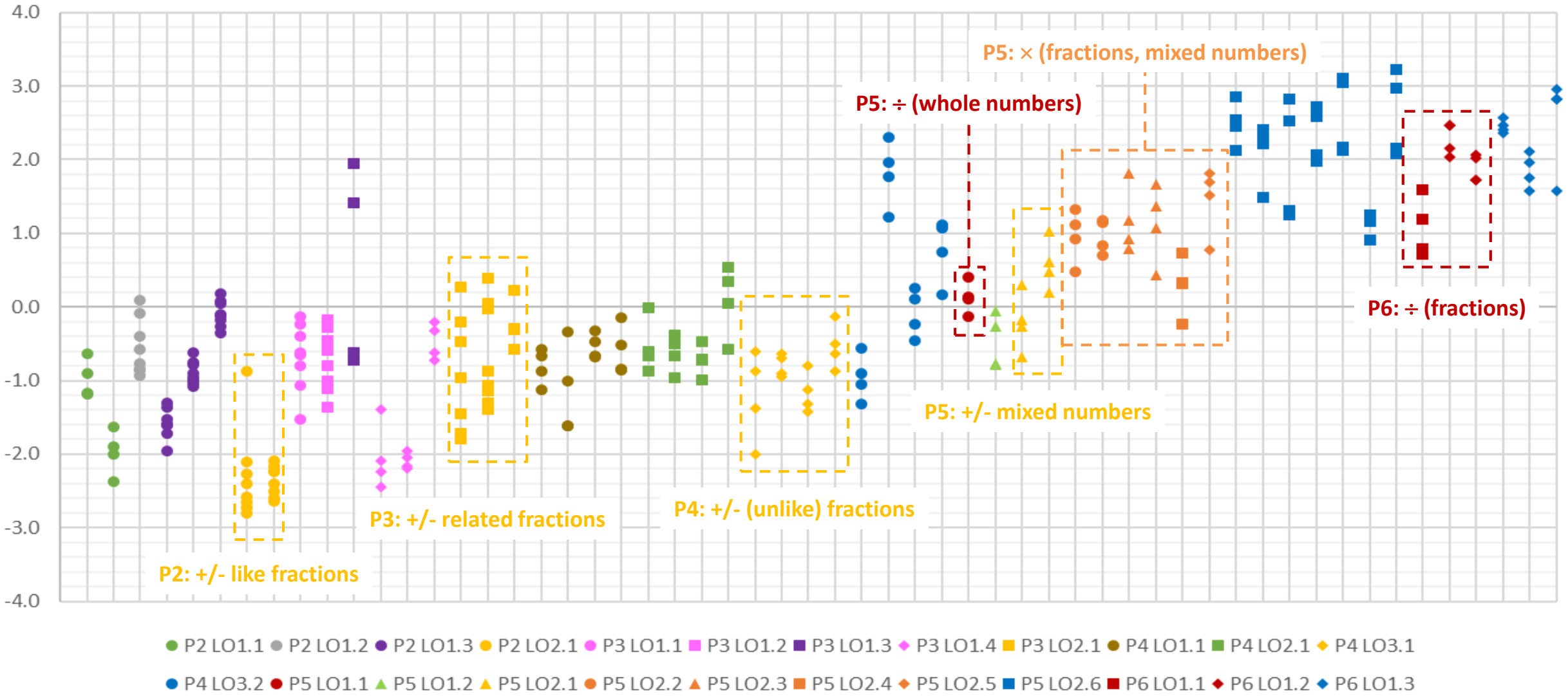


CAT Fractions Item Locations (Rasch measures)



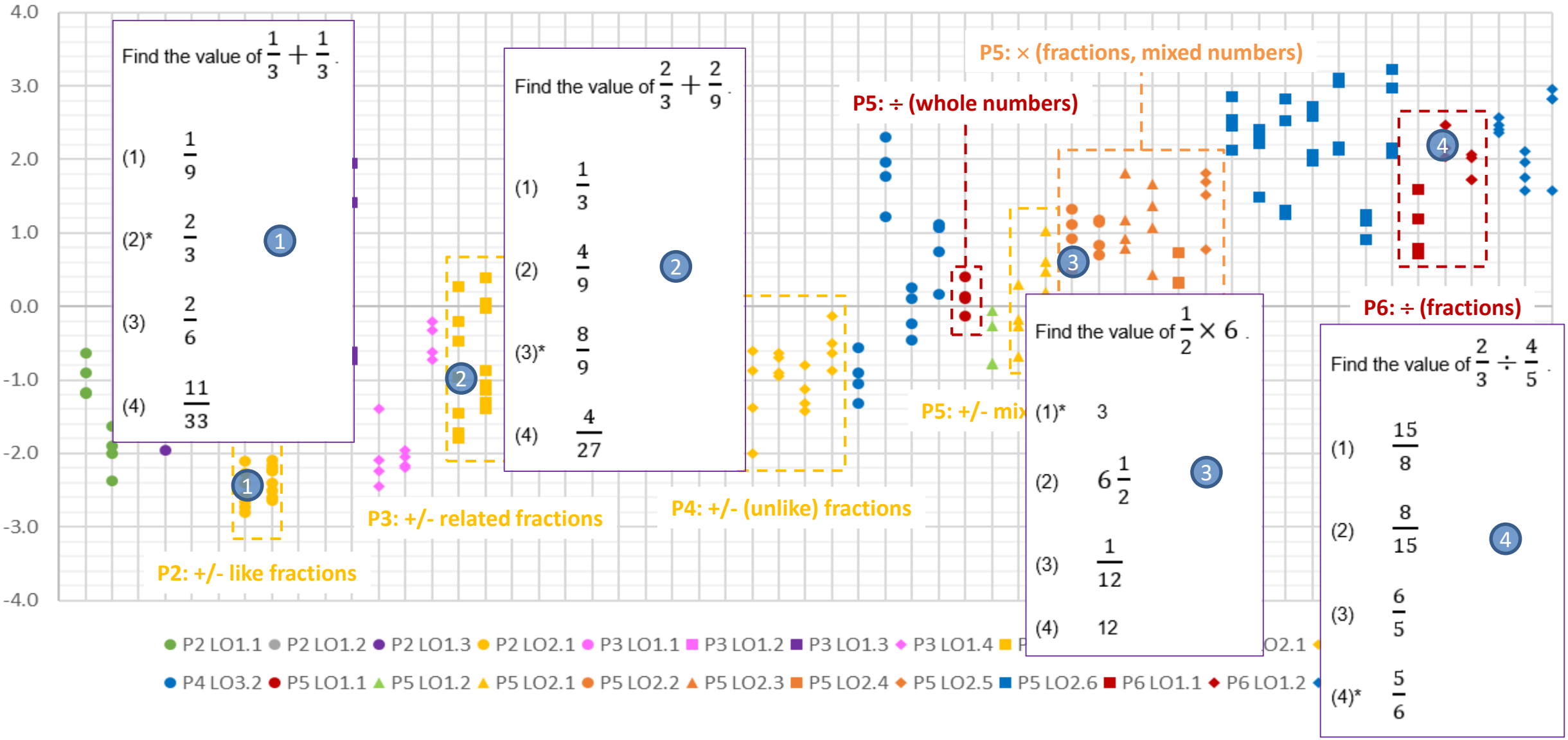
Items in each column are generated from the same item template (264 items from 56 item templates)

CAT Fractions Item Locations (Rasch measures) – relation with LOs

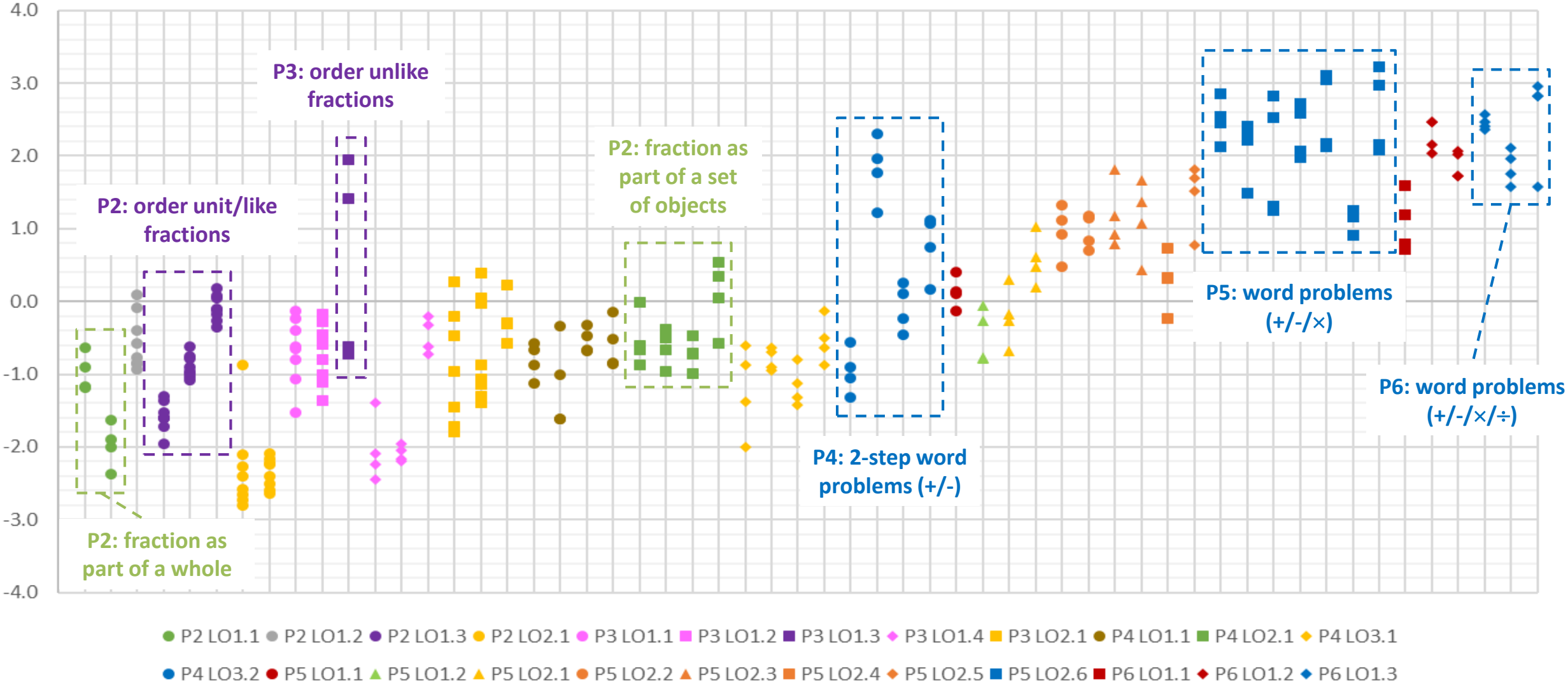


Items in each column are generated from the same item template (264 items from 56 item templates)

CAT Fractions Item Locations (Rasch measures) – sample items

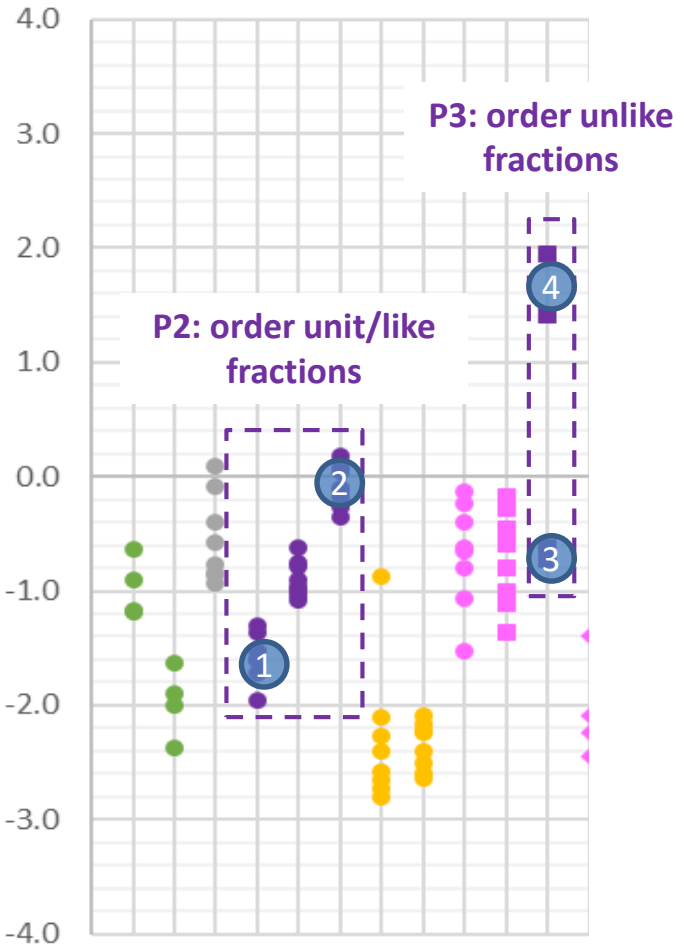


CAT Fractions Item Locations (Rasch measures) – relation with LOs



Items in each column are generated from the same item template (264 items from 56 item templates)

CAT Fractions Item Locations (Rasch measures) – sample items



Which one of the following fractions is the **smallest**?

(1) $\frac{1}{3}$

(2) $\frac{2}{3}$

(3)* $\frac{1}{12}$

(4) $\frac{11}{12}$

2

Which one of the following fractions is the **smallest**?

(1) $\frac{1}{2}$

(2)* $\frac{1}{5}$

(3) $\frac{3}{5}$

(4) $\frac{3}{10}$

4

Which one of the following fractions is the **smallest**?

(1) $\frac{1}{5}$

(2) $\frac{1}{6}$

(3) $\frac{1}{7}$

(4)* $\frac{1}{8}$

1

Which one of the following fractions is the **smallest**?

(1) $\frac{1}{3}$

(2) $\frac{1}{4}$

(3) $\frac{2}{4}$

(4)* $\frac{2}{12}$

3

Summary of Empirical Findings

- Good spread of item difficulties
- Reasonably matched to student abilities
  - Items are designed to assess students' mastery of learning objectives across different levels of primary school
- Sensible ordering of item/template locations by learning objectives
  - A more demanding LO, on average, gives rise to more difficult items/templates
  - *First order* impact on item/template difficulties
- The use of item templates facilitates the study of how various factors may impact on item locations/difficulties
  - E.g., specific steps/methods required to solve the problem, numerical values that give rise to different degree of complexity in arithmetic computation
  - *Second order* impact on item/template difficulties
  - Better understanding will enhance the control of item difficulties through the design of item templates

Conclusion

Achievements & Benefits

- 2 good quality item pools for CAT Fractions and CAT Decimals
  - 156 item templates, 735 items
- Optimised existing resources
  - Past exam items, MS Word, SAS
- Greater efficiency
- Deepen understanding on item construct and item difficulty

Further Work

- Improve techniques in developing item templates
  - To better control item difficulty and reduce need for calibration
- Develop software for automatic item generation

Singapore Examinations and Assessment Board

A trusted authority in examinations and assessment,
recognised locally and internationally

Acknowledgement

The following officers contributed in the exploration of AIG:

Geak Seng CHEN-THENG, Yick Chee FONG, Yuxin LIN, Andy LUO, Ai Lingg ONG, Jean PHUA