*Original Research Article*

Asian Journal for Mathematics Education
1-29

DOI: 10.1177/27527263221143802
journals.sagepub.com/home/mea

# Rational numbers and proportional reasoning in Chinese primary schools: Patterns, latent classes, and reasoning processes

**Zhenzhen Miao[1], Christian Bokhove[2], David Reynolds[3], and Charalambos Y. Charalambous[4]**

**Abstract**

Rational numbers and proportional reasoning are challenging topics in school mathematics. Though much research has been conducted, the learning of these topics is rarely explored among learners taught by expert teachers. Existing research also lacks sufficient attention toward latent classes and reasoning processes among primary learners at scale. The current study thus looks into the learning of the topics among 2,642 fifth and sixth graders taught by master mathematics teachers in China. Utilizing items identified as top level tasks in previous research, the article captures performance patterns, latent classes and reasoning processes of the learners, from a cross-sectional perspective. Data are analyzed with Rasch, mixture Rasch model (MRM), and complementary statistical methods. Variation between grades and gender is scrutinized where possible. The results show that, at the macro level, students have an average success rate of 78%, 78%, and 52% in decimals, fractions, and proportional reasoning, respectively; sixth graders outperform fifth graders; gender gap exists in the majority of Grade 5 but closes in Grade 6. At the meso level, two to three latent classes emerge in each grade with differential performance. Microlevel analyses show that different characteristics of reasoning lead to different performance of members affiliated to different latent classes.

**Keywords**

mixture Rasch model, primary students, proportional reasoning, Rasch model, rational numbers

Date received: 16 August 2022; accepted: 21 November 2022

[1]School of Education, Jiangxi Normal University, Nanchang, China
[2]Southampton Education School, University of Southampton, Southampton, UK
[3]School of Education, Hangzhou Normal University, Hangzhou, China
[4]Department of Education, University of Cyprus, Nicosia, Cyprus

**Corresponding Author:**
Zhenzhen Miao, School of Education, Jiangxi Normal University, 99 Ziyang Ave, Nanchang 330022, China.
Emails: z.miao@jxnu.edu.cn; zhenzhenmiao@outlook.com

# 1. Introduction

Among topics of school mathematics, rational numbers and proportional reasoning are found to play a profound role in and pose a lasting effect on the mastery of algebra and overall mathematics performance in secondary and undergraduate phases (Hurst & Cordes, 2018; Siegler et al., 2012). It has been argued that these topics are poorly learnt and urgently in need of improvement even in the secondary grades (Brown et al., 2010; Hurst & Cordes, 2018; National Mathematics Advisory Panel, 2008; Torbeyns et al., 2015).

Internationally, the learning of rational numbers, ratios, and proportions is at the core of expectation and assessment in school mathematics. Regarding rational numbers, on the TIMSS 2015 international benchmark for Grade 4, advanced performers (the top) were expected to have "an increasing understanding of fractions and decimals"; and high performers (next to the top) to apply "simple fractions and two-place decimals" in word problems (Mullis et al., 2016, Exhibit 2.1). On the eighth grade benchmark, top performers were expected to tackle various problems involving fractions, proportions and percentages and "justify their conclusions"; high performers to build connections between fractions, decimals, ratios, and percentages and "solve problems with fractions, proportions, and percentages" (Mullis et al., 2016, Exhibit 2.13). With similar expectations in TIMSS 2019, fourth graders had reached an average correct rate of 47% across participating countries/economies on a word problem demanding the subtraction of a non-unit fraction from 1 (Mullis et al., 2020, Exhibit 1.12.2). The international correct rate dropped to an average of 18% for eighth graders on a multistep problem that was embedded in a $3 \times 3$ sudoku and involved subtracting two fractions with unlike denominators from 1 for at least twice: $1 - \frac{2}{5} - \left(1 - \frac{1}{5} - \frac{8}{15}\right)$ (Mullis et al., 2020, p. 192, Exhibit 3.13.1).

In assessments focusing on higher-order mathematics competencies, rational numbers and proportional reasoning also play a major role in defining the proficiency of certain levels, such as the Level 3 of mathematics in the Programme for International Student Assessment (PISA) (OECD, 2019, p. 105, Table I.6.1). In PISA 2018, 53.8% of all participants managed to reach the mathematics Level 3, while over 90% of the 15-year olds from the four Chinese provinces claimed the proficiency on this level (OECD, 2019). In fact, like in many countries, in China, students start to learn these topics in primary schools. More specifically, the National Curriculum in China (Ministry of Education China, 2011, mandated version at the time of the study) expects students in the upper-primary grades (i.e., Grades 4–6) to (1) understand the meanings of decimals, fractions, and percentages and make conversions and comparisons between them; (2) add, subtract, multiply, and divide decimals and fractions; (3) do mixed operations with decimals and fractions in mainly two steps but no more than three steps; (4) solve real-world problems involving decimals and fractions (including percentages); (5) understand the meanings of ratio and proportionality and be able to solve simple problems; (6) decide if two quantities in a real-world context are directly or inversely proportional; (7) be able to graph directly proportional data and estimate a missing value based on a given value from the data; and (8) find direct and inverse proportional examples from the real world and talk about them with others. The *expected curriculum* is after all not the *obtained curriculum* (Valverde et al., 2002). In the primary phase, there has been recent evidence on Chinese students' deficiency in decimals and fractions. Chinese students were found yet to overcome the influence of the discreteness of whole numbers on their perception of decimals as a continuous quantity even by the end of primary schooling (Liu et al., 2014). From mixed schooling backgrounds in Beijing, 310 sixth graders were found able to tackle part-whole problems but inadequate in solving problems that involved repetitive partitioning and location of fractions on the number line (Jiang et al., 2021). Albeit recent evidence on performance in decimals and fractions, there is a dearth of empirical evidence on Chinese students' performance in rational numbers and proportional reasoning as a combined focus, given the connectedness between the topics suggested by existing work.

There is also a lack of research focusing on students taught by expert teachers, for example, master teachers, in China. Studying students taught by master mathematics teachers may help expand our understanding of what is possible for every child, should excellent teaching be provided, given that expert teachers have been theorized and found to be able to best implement the curriculum and nurture excellent learning outcomes (Berliner, 2001; Leinhardt, 1989) in countries other than China (see Bond et al., 2000, for evidence from the United States). Researching on the performance of students in master teachers' classes is also a necessary response from the educational research sector to the educational practice sector in China where master teachers play a leading role in teacher professional development (PD) (Cravens & Wang, 2017; Fan et al., 2015; M. Zhang et al., 2021). In China, teachers are promoted, through the official rank system, ascending from third level to second level, first level, Senior, and then Professoriate Senior. Beyond the five official ranks are honorary tiles, such as Super Teachers, Subject Leaders, and Backbone Teachers, recognizing teachers for excellent teaching practice and strong contribution to peers' PD, in addition to other merits (Fan et al., 2015; Huang et al., 2017; Li et al., 2011). Within provinces, there are funding opportunities for master teachers to apply for Master Teacher Studios and, upon approval, to carry out cross-school PD projects over a period of 2 to 3 years, in addition to their contribution to the teaching research group (TRG)-based within-school PD (Li et al., 2011; M. Zhang et al., 2021). While existing research has tapped on master teachers' practices, PD, and beliefs, often through relatively small-scale studies, little has been explored at scale with respect to their students' learning performance.

More empirical evidence is also needed on the variation of performance within learners at scale—particularly evidence generated with a model that considers both examinee clustering and psychometric measurement. The mixture Rasch model (MRM) is the one that does both. MRM was used in research on *teachers*' competence in and understanding of rational numbers and proportional reasoning (Izsák et al., 2010; Ölmez & Izsák, 2021). Yet, existing research lacks empirical insights into the topics using similar approaches to data from primary school *learners*, particularly those taught by expert teachers.

In addition to filling the gaps discussed above, the study aims to contribute to research on the learning of fractions, decimals and proportional reasoning by studying the learning of the topics together. The sample consisting of Grades 5 and 6 also makes it possible for the study to estimate learning trends chronologically between grades.

As a part of a large-scale project on mathematics learning and teaching in master mathematics teachers' classes, the current study aims to (1) uncover performance patterns of fifth and sixth graders taught by master teachers on rational numbers and proportional reasoning using the Rasch model (macro level); (2) identify latent classes among the children with the MRM (meso level); and (3) track and diagnose cognitive "footsteps" that learners in different latent classes leave behind as they strive to reason rationally and proportionally in problem solving (microlevel).

## 2. Theoretical framework

### *2.1 Rational number sense and proportional reasoning*

The concept of rational numbers features "a set of related but distinct subconstructs", commonly known as measure, quotient, operator, ratio, and part-whole (Behr et al., 1983). At the core of the rational number construct and subconstructs are two elements of one unified scheme: the concept of unit and the process of partitioning (Carpenter et al., 1993/2009). They form the basis for two aspects of rational numbers—rational numbers as *quantities* and rational numbers as *relationships* between quantities (Nunes et al., 2009). Mastery of these two aspects in turn leads to optimal *rational number sense* and *proportional reasoning*. The whole can be infinitely equally partitioned into smaller units, hence the existence of an infinite number of numbers between two seemingly adjacent numbers, say 0.123 and 0.124, or 1/123 and 2/123. On the one hand, infinite partitioning suits the need for number operation or number size perception or comparison. It makes rational numbers continuous, in contrast to the discreteness of whole numbers. On the other hand, it generates an infinite set of equivalents for any given fraction or ratio, allowing for reasoning about two varying quantities that define a fraction or ratio on the ground of an invariant relationship between the two quantities (Charalambous & Pitta-Pantazi, 2005, 2007).

The subconstructs of rational numbers are fundamentally connected (Carpenter et al., 1993/2009; Empson et al., 2011; Freudenthal, 1983), albeit differing *phenomenologically* in contexts where they are situated and applied (Freudenthal, 1983). As a phenomenon (or subconstruct) of rational numbers, ratio can be about magnitudes from the same measurement space or different measurement spaces, with the former representing within ratios and the latter between (Freudenthal, 1973, 1978). For example, for one of the "eel-feeding" items developed by Piaget et al. (1968, 1977), within and between ratios can be expressed as shown in Figure 1. In fact, there are two constant loops running through within and between ratios in a given proportional situation where one could reason within or between two measurement spaces in the direction of clockwise or counter clockwise or diagonally (Figure 2). The latter turns it into an inverse proportional model where the product of two numbers on one diagonal is equal to that on the other.

Research on the topics has also been approached from the perspectives of curricula (Y. Li et al., 2009; Moseley et al., 2007; Yang et al., 2010) and classroom teaching (Cai & Wang, 2006), with endeavors from both sides identifying an emphasis on connections between division, fractions, and ratios. On the one hand, division, fractions, and ratios all represent relationships between two quantities: the divisor versus the dividend, the numerator versus the denominator, and the first term versus the second term. On the other hand, they each result in a value that shows the strength of the relation. What connects the three concepts at a deeper level is the similar property—multiplying or dividing both terms of a fraction, ratio, or division by a nonzero number, the value of the fraction, ratio, or quotient remains unchanged: (1) $\frac{a}{b} = \frac{a\times n}{b\times n} = \frac{a\div n}{b\div n}$; (2) $a{:}b = (a\times n){:}(b\times n) = (a\div n){:}(b\div n)$; (3) $a\div b = (a\times n)\div(b\times n) = (a\div n)\div(b\div n)$ ($b\neq 0$ and $n\neq 0$).

Following such fundamental connections, direct proportionality is essentially about equivalent ratios, fractions, or quotients. These all can be expressed by the function $y=f(x)=kx$, where $k$ is a constant (Piaget et al., 1977). In the real world, the label for the constant $k$ is "situation-specific" which can be

**Within ratios**

| | | | | | | |
|---|---|---|---|---|---|---|
| 5cm | : | 10cm | = | 1 | : | ? |
| *length of eel A* | | *length of eel B* | | *n of pellet fed to eel A* | | *n of pellet fed to eel B* |

**Between ratios**

| | | | | | | |
|---|---|---|---|---|---|---|
| 5cm | : | 1 | = | 10cm | : | ? |
| *length of eel A* | | *n of pellet fed to eel A* | | *length of eel B* | | *n of pellet fed to eel B* |

Source: Freudenthal (1973, 1978); Piaget et al. (1977).

**Figure 1.** Within and between ratios.
*Source*: Freudenthal (1973, 1978); Piaget et al. (1977).

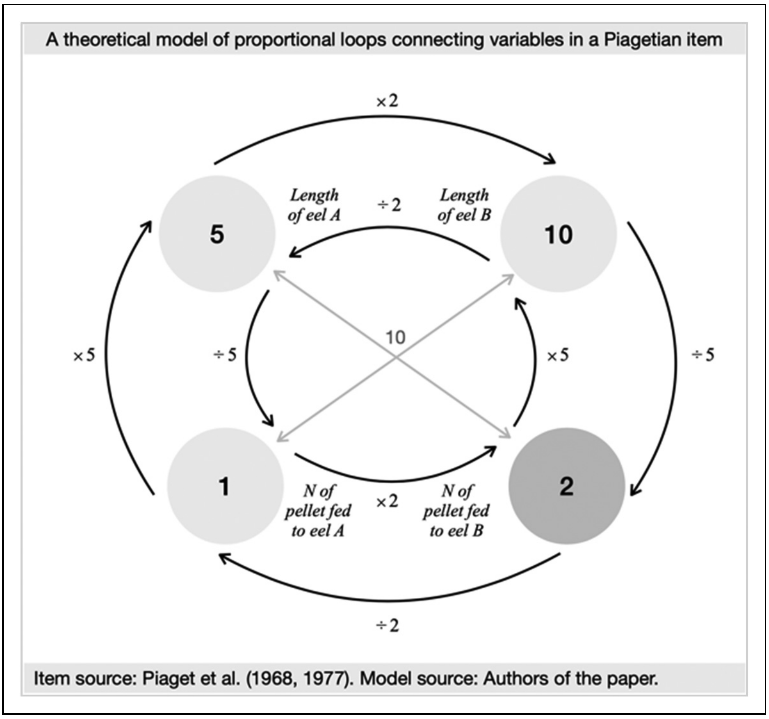



**Figure 2.** A theoretical model of the constant multiplicative loops in proportionality.

the constant rate of change, the scale of a map, the scale factor for shrinking/enlarging, a percentage about tax, or a theoretical probability in dice rolling (Lamon, 2007, p. 639). It is crucial that in situations underpinned by this function, learners choose multiplicative over additive strategies and reason proportionally about variant quantities constrained by this invariant ratio ($k = \frac{y}{x},\ x \neq 0$) as they strive to reach a fruitful solution (Piaget et al., 1977). Proportional reasoning is ultimately about a function that defines invariant relationships between variables. What is invariant could be a ratio in direct proportionality or a product in inverse proportionality. An indicator of "rational number sense", proportional reasoning is gradually developed over the course as a learner experiences a rich collection of traits of rational numbers (Lamon, 2007, p. 640). In time, learners evolve from thinking around whole numbers in the primary grades to reasoning about rational numbers in the secondary (Hiebert & Behr, 1988). On this "evolving" path lurk potential obstacles in the form of various misconceptions.

## *2.2 Misconceptions in rational number sense and proportional reasoning*

Misconceptions about the topics have received much attention from research which tends to generate general patterns of learners' errors. The existing evidence suggests in general the infinite nature of rational numbers appears to be challenging to learners whose perceptions are hindered by their knowledge of the discreteness of whole numbers (Hart et al., 1985b; Liu et al., 2014). Natural number bias often leads to errors in rational number operations when items are incongruent (González-Forte et al., 2021). Similarly, Chinese sixth graders from mixed schooling backgrounds found partitioning and locating fractions on the number line the most challenging (Jiang et al., 2021).

Errors in solving fraction tasks, particularly challenging tasks, might be rooted in incompetence in locating the right referent unit, inability of recognizing when and how to represent parts of parts of a whole accurately (Izsák et al., 2010) and inaccurate estimation of fraction magnitude

(Copur-Gencturk, 2021) among teachers as well as the lack of knowledge of various representations of fractions among students (Hart, 1987). Teachers were found to differ considerably in their understanding of the arithmetic rationale behind fraction division (Ma, 1999) and their capacity in making sense of learner's misconceptions in fractions (Charalambous, 2016). Their development in the teaching of the topics led to the mastery of these topics among their students even including those immensely struggling at mathematics (Howe et al., 2015; S. Zhang et al., 2021).

In proportional reasoning, errors are found largely due to the tendency of choosing additive over multiplicative strategies (Hart, 1980; Misailidou & Williams, 2003)—a dilemma that students often encounter in problem situations underpinned by proportionality (Thompson & Thompson, 1994; Van Dooren et al., 2010). Those taking multiplicative strategies might still end up with a wrong solution because simple doubling or halving was carried out to tackle ratios that were not as straightforward as 1:2 or 2:1 (Hart, 1980).

## 3. Research questions (RQs)

As part of a larger project, the study sets out to explore patterns, latent classes and reasoning processes of rational numbers and proportional reasoning of primary fifth and sixth graders ($n = 2{,}642$) taught by master mathematics teachers in urban China. The current article seeks to address three RQs:

RQ1. How do the students taught by master mathematics teachers perform in rational numbers and proportional reasoning?
RQ2. Among the participants, are there distinct latent classes that perform significantly differently on rational numbers and proportional reasoning?
RQ3. What are the typical reasoning processes generated on the items that strongly distinguish latent classes?

To address the RQs, the Rasch model, MRM, and in-depth diagnoses of reasoning processes were conducted, with complimentary statistics assisting comparisons between groups. The following section details the research design and methods.

## 4. Methods

### 4.1 Participants

Piloted in Spring and Autumn of 2018, the larger project recruited master mathematics teachers and their students, through the combination of stratified sampling and expert recommendation which located cities first and then teachers and their classes, with one teacher corresponding to one class. The project was approved by the first author's institution, and informed consents were obtained before data collection. Three major steps were taken in obtaining a final sample for the project:

First, according to the Gross Domestic Product (GDP) per capita ranking of major cities in China in 2018 (National Bureau of Statistics of China, 2019), we located the sampling sites—capital cities and economically equivalent cities in five Chinese provinces/municipalities, Anhui, Beijing, Jiangsu, Jiangxi, and Tianjin. These cities spread about evenly across the first, second and third quarters of the 2018 ranking.

Second, the provincial Teaching Research Officials (TROs) responsible for primary mathematics, as gatekeepers of mathematics teaching and learning in their home provinces, were asked to help recruit around 15 master mathematics teachers who demonstrated best teaching practice in primary mathematics and were full-time classroom teachers teaching mathematics to primary grades, ideally Grades 5 and 6. The inclusion criteria seek to look for master teachers who demonstrate best primary mathematics teaching in their home province and (1) have the ranks of Professoriate Senior or Senior; (2) are recognized as Super Teachers, Subject Leaders, or Backbone Teachers; and (3) have won teaching awards in provincial and/or national teaching competitions.

Based on the TROs' recommendations, we had an initial sample of 81 teachers and 3,737 students. With informed consent from participants, we collected the initial bulk of data in the 81 teachers' classes, including lesson observations. To make sure that the teaching data collected represented daily teaching that would happen the same way without the appearance of external observers and cameras, we applied one more criterion (the third and last step) to screen and finalize the sample based on the authenticity of the lesson data, though every effort had been taken to minimize observer effects during the data collection. The criterion was simple: the lesson must be a typical normal lesson that was unfolded as usual after its precedent lesson. This resulted in 70 teachers' lessons meeting the criterion. We had to drop data related to 11 teachers and their students because their lessons were carefully prepared teaching research lessons (教研课) which did not represent the kind of everyday teaching we were explicitly aiming to observe.

The final sample of the project (Table 1) included 3,178 students in Grades 2–6 and 70 master mathematics teachers of which 61 were teaching Grades 5 or 6 in the 2018–2019 academic year. All 70 teacher participants (47 females) are full-time mathematics teachers who are in their 30s ($n = 24$), 40s ($n = 43$), or 50s ($n = 3$). The majority ($n = 46$) of the teachers have 20 to 29 years' mathematics teaching experience; 20 teachers have 10 to 19 years' experience; and 4 have 30 to 38 years' experience. In terms of professional ranks, 18 of the teachers are First level, 43 are Senior, and 9 are Professoriate-Senior. In terms of honorary titles, 17 are Backbone Teachers, 40 teachers are Subject Leaders, and 13 are Super Teachers. They have all won multiple teaching awards in teaching competitions at provincial and national levels. In addition to their teaching obligations, 48 of them are also (deputy) head teachers or (deputy) heads of the TRG for mathematics at their home schools.

Multiple types of data were collected in Spring 2019. As part of this project, a mathematics test was administered to 2,642 students in Grades 5 and 6 of which 1,318 were fifth graders ($11.1 \pm 0.4$ years), and 1,324 sixth graders ($12.0 \pm 0.4$ years), including 1,204 girls, 1,433 boys, and five students who completed the test but did not give gender information. This article focuses on students' performance in the mathematics test.

As follows, we explain the source, structure, and expectation of the items utilized in the test.

## 4.2 The test

To assess students' performance in rational numbers and proportional reasoning, we utilized a set of top level items (Table A1) of *Decimals*, *Fractions*, and *Ratio and Proportion* from the classical Concept in Secondary Mathematics and Science (CSMS) project in England (Hart et al., 1981). The tests developed on the CSMS project were later published as the Chelsea Diagnostic Tests, with a Teacher's Guide providing detailed grading criteria (Hart et al., 1985a, 1985b). In 2009, a set of same items were utilized on the Improving Competence and Confidence in Algebra and Multiplicative Structures (ICCAMS) project (Hodgen et al., 2010) which found a considerable decline of performance, in comparison with the CSMS results, in rational numbers, proportional reasoning, and algebraic thinking among a nationally representative cohort of English secondary school students 30 years on. Utilizing the classical items helps gain insights into variation of mathematical proficiency and reasoning patterns between cohorts of students, albeit the necessity to consider potential differences between the underlying populations.

*4.2.1 Decimals.* In the CSMS project, there were six levels in decimals (Hart et al., 1985b). In the current study, fifth graders attempted decimal items at Levels 5 and 6. The level proficiency thresholds for decimals were a correct rate of 2/3 on Level-5 items and 3/4 on Level-6 items. These top level items demand:

- a fundamental understanding of the infinite feature of rational numbers (e.g., item D03);

**Table 1.** Demographic information of master teachers and their students.

| | Students | | | Teachers | | |
|---|---|---|---|---|---|---|
| | *N* | *Age* | % | *N* | *Age* | % |
| *Province* | *3,033* | *11.3 (0.9)* | *100%* | *70* | *41.5 (4.4)* | *100%* |
| Anhui | 673 | 11.0 (0.9) | 22.2% | 14 | 43.0 (5.6) | 20.0% |
| Beijing | 523 | 11.5 (0.5) | 17.2% | 16 | 39.0 (3.6) | 22.9% |
| Jiangsu | 655 | 11.2 (1.1) | 21.6% | 15 | 42.4 (3.5) | 21.4% |
| Jiangxi | 733 | 11.2 (1.2) | 24.2% | 14 | 41.4 (4.1) | 20.0% |
| Tianjin | 449 | 11.7 (0.6) | 14.8% | 11 | 42.8 (4.3) | 15.7% |
| *Grade* | *3,033* | | | *70* | | |
| 2 | 54 | 8.1 (0.4) | 1.8% | 2 | 38.5 (3.5) | 2.9% |
| 3 | 134 | 9.3 (0.4) | 4.4% | 2 | 45.5 (0.7) | 2.9% |
| 4 | 203 | 10.0 (0.6) | 6.7% | 5 | 45.0 (4.5) | 7.1% |
| 5 | 1,318 | 11.1 (0.4) | 43.5% | 32 | 41.2 (4.5) | 45.7% |
| 6 | 1,324 | 12.1 (0.4) | 43.7% | 29 | 41.3 (4.2) | 41.4% |
| *Gender* | *3,028* | | | *70* | | |
| Female | 1,392 | 11.2 (1.0) | 46.0% | 47 | 41.3 (3.9) | 67.1% |
| Male | 1,636 | 11.3 (0.9) | 54.0% | 23 | 42.0 (5.3) | 32.9% |
| *Teacher position within schools* | | | | *70* | | |
| Teacher with no other obligations | | | | 22 | 42.1 (5.9) | 31.40% |
| Head or Deputy Head of Maths TRG | | | | 32 | 40.3 (2.9) | 45.70% |
| Head teacher or Deputy Head teacher | | | | 16 | 43.1 (4.1) | 22.90% |
| *Teacher rank* | | | | *70* | | |
| First level (一级) | | | | 18 | 39.9 (3.7) | 25.7% |
| Senior (高级) | | | | 43 | 41.8 (4.5) | 61.4% |
| Professoriate Senior (正高级) | | | | 9 | 43.7 (4.5) | 12.9% |
| *Honorary titles* | | | | *70* | | |
| Backbone Teacher (骨干教师) | | | | 17 | 39.0 (4.0) | 24.3% |
| Subject Leader (学科带头人) | | | | 40 | 41.6 (3.4) | 57.1% |
| Super Teacher (特级教师) | | | | 13 | 44.7 (5.6) | 18.6% |
| *Mathematics teaching experience* | | | | *70* | | |
| 10–19 years | | | | 20 | 37.1 (2.0) | 28.6% |
| 20–29 years | | | | 46 | 42.6 (2.8) | 65.7% |
| 30–38 years | | | | 4 | 51.5 (4.0) | 5.7% |

*Note. TRG* = teaching research group (教研组). The project's student sample consists of 3,178 students of which all were in the lessons observed, 3,033 completed a questionnaire and provided their demographic information in the course of completing the questionnaire, and 145 students did not complete the questionnaire because of scheduling difficulties. The student information presented in the table comes from the 3,033 students.

- the proficiency in calculating or estimating division answers that are not integers but decimals, no matter the dividend is greater (item D04) or smaller (items D05 and D06) than the divisor;
- the fluency in making conversion between different decimal places, for example, tenths to hundredths (item D02); and
- the ability to write a decimal according to the verbal expression of its place value, particularly when it is expressed in such a way that the number of units on the place exceeds 10, for example, writing 11 tenths as 1.1 (item D01).

*4.2.2 Fractions.* In the CSMS, there were two fraction tests (Hart et al., 1985b), Fractions 1 (F1) and Fractions 2 (F2), each consisting of items situated at four hierarchical levels. The current study included items of Level 4 from F1 (F1-L4) and Levels 3 and 4 from F2 (F2-L3; F2-L4) which expected students to:

- apply the knowledge of equivalent fractions (items F05, F06, F07, F08, and F09);
- model and solve real-world problems using subtraction of fractions with different denominators (item F03);
- model and solve real-world problems by dividing fractions that differ in both numerators and denominators (item F10);
- model and solve real-world problems by dividing two mixed fractions (item F04); and
- reason abstractly about fractions in an algebraic situation and provide the necessary condition that satisfies the situation (e.g., $\frac{a}{b} < \frac{a}{c}$, in item F11).

The level proficiency thresholds for fractions were 3/4 correct on F1-L4 items, 2/3 correct on F2-L3 items, and 3/5 correct on F2-L4 items.

*4.2.3 Ratio and proportion.* In the CSMS study's ratio and proportion test (Hart et al., 1985b), there were four levels. In the current study, both fifth and sixth graders attempted the Level-3 items, whereas the latter also took the Level-4 items. The level proficiency thresholds for ratio and proportion were a success rate of 2/3 on Level-3 items and a success rate of 3/4 on Level-4 items. The top level items of ratio and proportion demand:

- the ability to identify and apply ratios that are much harder than 2:1 or 1:2 and are set in real-world contexts, for example, where two terms of a ratio is enlarged or shrunk by a fraction (rather than a whole number) with its numerator and denominator being coprime but not 1 (e.g., 9/8 in item R09) or by a fraction that can be simplified as such (e.g., 12/18 in item R10);
- the ability to reason proportionally, use ratio to model real-world problems, and ultimately solve the problems (e.g., items R02, R03, R04, R06, and R10); and
- the ability to model and solve real-world problems in more complex situations where three or more quantities are interrelated proportionally (e.g., items R02, R03, R04, and R08).

The CSMS items were translated, back translated, validated, and carefully selected after a pilot in one Grade 5 and two Grade 6 classes in a participating school. The final set of 28 items (Table A1) were put into three booklets, G5A, G5B, and G6, each consisting of 12 to 16 items. There were seven common items (items F01, F11, R01, R02, R03, R04, and R06) shared between the three booklets for test-equating purposes.

## *4.3 Data collection*

The mathematics test was conducted in the form of paper and pencil in a usual school slot of 40 min. Students were given one of the three test booklets. Three hundred and seventy-nine fifth graders in Anhui took the booklet G5A, 939 fifth graders in other four provinces took the booklet G5B, and 1,324 sixth graders from all five provinces took the booklet G6.

## *4.4 Data analyses*

Data analytic strategies were of three folds capturing macro, meso, and micro patterns of learners' cognitive status in responding to the test items. Rasch analyses show the macro pattern of mathematics performance; MRM captures the variation of the assessed competence through the lens of latent classes at the meso level; and cognitive diagnoses put together the micro detail of various reasoning processes that students go through and shed light on key features of cognition that distinguish one latent class from another. Descriptive and comparative analytical approaches were applied in addressing the three RQs.

*4.4.1 Data coding.* The test data were coded with the CSMS marking schemes (Hart et al., 1985b) which provided both dichotomous and multi-category codes for each item. Dichotomous coding (1 = correct; 0 = wrong) prepared data for psychometric analyses, Rasch and mixture Rasch, which in turn offered us macro and meso patterns of learning. Multi-category coding categorizes in detail various responses to each test item, making it possible to diagnose learners' reasoning processes at the microlevel. Take the item F05 (Table A1) as an example for the multi-category coding: a code 1 was given when 2/3 of the given shape was shaded; a code 6 was given when a half was shaded; a code 7 was given when 1/3 was shaded; a code 9 was given when other shading strategies were used; and a code 0 was given if blank. As the result of a two-day training, adequate inter-coder agreement (pairwise) had been reached (Cohen's kappa: $\bar{\kappa}_1 = .95$, $SD_{\kappa_1} = .08$; $\bar{\kappa}_2 = 1.00$, $SD_{\kappa_2} = .00$) through two rounds of ratings on a total of 22 test papers by the first author (master coder) and three undergraduates majoring in primary mathematics education. With inter-rater agreement established, the three paid assistants went on coding the rest of the test papers.

*4.4.2 Macro analyses: Estimating intended competence with Rasch model and test linking.* To capture the grand pattern of learners' proficiency, the Rasch analysis was carried out with the weighted likelihood estimation method in the R package, TAM (Kiefer et al., 2021). The weighted mean square (infit MNSQ) values ($M = 0.99$, $SD = 0.1$) were all within the interval of [0.7, 1.3], indicating an adequate model fit (Bond & Fox, 2015). To link the three tests, person and item parameters in G5A and G6 were shifted onto the same scale as of G5B, with the seven common items acting as link/anchor items, using the equations below (Wu et al., 2016):

$$T'_{G5A} = \frac{(T_{G5A} - \mu_{G5A})}{\sigma_{G5A}} \sigma_{G5B} + \mu_{G5B}, \tag{1}$$

$$T'_{G6} = \frac{(T_{G6} - \mu_{G6})}{\sigma_{G6}} \sigma_{G5B} + \mu_{G5B}, \tag{2}$$

where $T'_{G5A}$ and $T'_{G6}$ were the shifted and scaled parameters for tests G5A or G6; $T_{G5B}$ and $T_{G6}$ were their initial parameters calibrated within the two tests, respectively; $\mu_{G5A}$, $\sigma_{G5A}$, $\mu_{G6}$, $\sigma_{G6}$, $\mu_{G5B}$, and $\sigma_{G5B}$ were the means and standard deviations (*SD*s) of the parameters of the seven link items in tests G5A, G6, and G5B, respectively. After the shift, person scores (Rasch logits) were linearly transformed into scores with an mean of 500 and an *SD* of 100.

*4.4.3 Meso analyses: Identifying latent classes through the MRM.* The MRM brings together two types of latent variable models, the latent class analysis and item response theory (Rost, 1990). The exploration of possible clustering within each test was conducted with the aid of the snowRMM module (Seol, 2020) in the *R*-based software, *jamovi* (The jamovi project, 2021), which adopts the joint maximum likelihood estimation method. Since not every case got a distinguishable estimate of class membership, the number of cases resulted in the MRM was smaller than the sample size. The optimal number of latent classes was decided through the comparison of the Bayesian Information Criterion (BIC) of each classification, as the BIC is found more reliable (F. Li et al., 2009). The number of latent classes with the lowest BIC was deemed appropriate.

*4.4.4 Micro analyses: Diagnosing reasoning processes across latent classes.* Multiple-category responses to each item were analyzed and compared between latent classes. Initially, chi-square tests of the distribution of various responses were carried out, with effect sizes (φ) calculated where a significant difference was observed. Then, we examined closely children's responses to the items that received a strong effect from class memberships. This involved the analyses and comparisons of proportions of each response within and between latent classes through pairwise $z$ tests for two proportions, using the equation 3:

$$z = \frac{(p_1 - p_2) - 0}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \tag{3}$$

where $p_1, p_2, n_1$, and $n_2$ were the proportions of the focal response and sizes of latent classes 1 and 2 respectively, $\hat{p}$ was the grand proportion of the response in the two classes under comparison as a whole. The calculation was carried out in Excel. The threshold for rejecting the null hypothesis that two proportions are equal is $p < .05$ ($|z| > 1.96$). A $p$-value smaller than .01 ($|z| > 2.58$) or .001 ($|z| > 3.29$) was also specified.

*4.4.5 Descriptive and comparative analyses.* To seek answers for RQ1, the overall patterns of mathematics performance at test, topic, and item levels were analyzed and compared between grades and gender ($t$-tests); item facilities (correct rates) and level proficiencies were calculated according to the CSMS standards (Hart et al., 1985b). Similar basic statistics were performed upon comparisons within and between latent classes, which addresses RQ2 and RQ3.

## 5. Results

In this section, we present and discuss the results of the measure of rational numbers and proportional reasoning at macro, meso, and microlevels, seeking to address the three RQs accordingly.

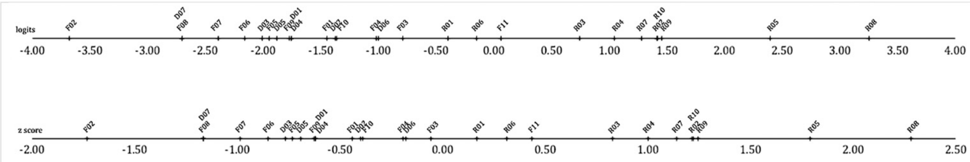
### *5.1 Macro patterns (RQ1): How do children taught by master mathematics teachers perform in rational numbers and proportional reasoning?*

*5.1.1 Overall performance.* After the shift and linear transformation of mathematics scores, mathematics performance among 2,642 children varied between 221 and 729. Unsurprisingly, sixth graders ($M = 541$, $SD = 109$) performed significantly better than fifth graders ($M = 458$, $SD = 68$) with a moderate effect, $t(2222) = 23.53$, $p < .001$, $d = .94$. Boys outperformed girls significantly with a weak to slightly modest effect, either across grades ($MD = 505 - 494 = 11$, $t(2635) = 2.85$, $p < .01$, $d = .11$) or within Grade 5 ($MD = 467–449 = 18$, $t(1306) = 4.58$, $p < .001$, $d = .27$), but not within Grade 6 ($MD = 545–539 = 6$, $t(1320) = 1.03$, $p = .30$).

*5.1.2 Item difficulties, facilities, and top level proficiencies.* After the test linking process, parameters ($M = -0.69$, $SD = 1.73$) of the 28 items ranged from −3.68 to 3.25 logits (Figure 3). The hardest were ratio items most of which were 1 $SD$ or farther above the mean. Students in both grades demonstrated a much higher proficiency on fractions and decimals than ratios, with the facility of either averaged to 78%. On 10 ratio items, fifth graders achieved an average facility of 38% ($SD = 14\%$), whereas sixth graders 59% ($SD = 12\%$).

Figure 4 shows facilities of all items that were given to fifth and/or sixth graders. Approximately 2/3 or more students succeeded on most decimal and fraction items. The items that appeared to be challenging were F11, R02, R03, and R04 for fifth graders and R05 and R08 for sixth graders (see Table A1 for a description of the items). In decimals, 71% of students taking G5A reached the top level (L6); 57% of those taking G5B reached Levels 5 or 6. In fractions, 61% fifth graders performed at Levels 3 or 4, with 55% of sixth graders reaching the top level, Level 4. In ratios and proportions, 16% of fifth graders and 48% of sixth graders reached Levels 3 or 4, whereas 42% of sixth graders also reached the top level, Level 4.

It remains unclear whether there exist latent classes of children performing considerably differently from each other and if so in what way. The next section seeks to address this question.




| Topic | *Mean facility* | *Mean* (logit) | *SD* (logit) | *Mean* (z) | *SD* (z) |
|---|---|---|---|---|---|
| Decimal | 78% | -2.07 | 0.88 | -0.80 | 0.51 |
| Fraction | 78% | -1.75 | 1.01 | -0.61 | 0.59 |
| Ratio | 52% | 0.83 | 1.37 | 0.88 | 0.79 |

**Figure 3.** Item parameters shifted on the same scale (logits and standardized z).

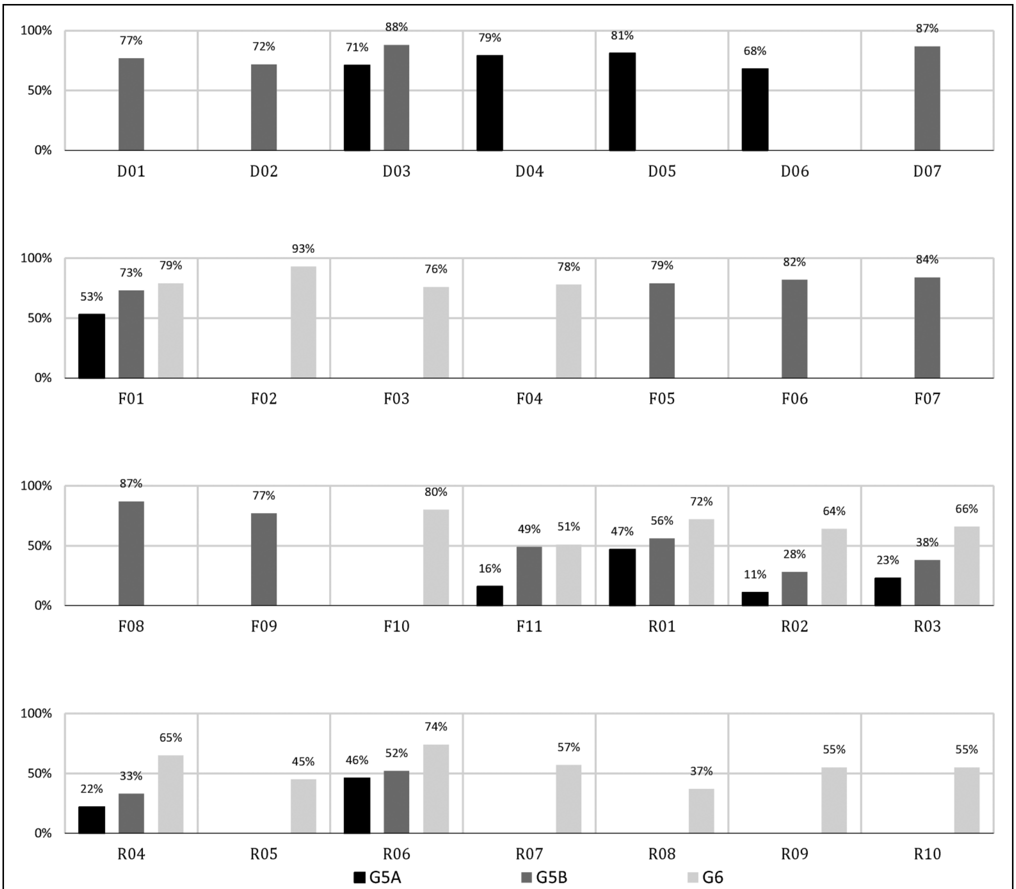



**Figure 4.** Item facilities.

### 5.2 Meso variation (RQ2): Among the participants, are there distinct latent classes that perform significantly differently on rational numbers and proportional reasoning?

Judging by the BIC values (i.e., the lower, the better) of the MRM, an optimal number of classes was identified within each of the three tests (Table 2). This resulted in the location of two latent classes in G5A, three in G5B, and three in G6. Most participants were allocated with latent class memberships, including 359 of 379 fifth graders in test G5A (94.7%), 829 of 939 fifth graders in test G5B (88.3%), and 1,151 of 1,324 sixth graders in test G6 (86.9%). Those who did not have a membership were (a) perfect (full) scorers (6 in G5A, 87 in G5B, and 126 in G6), (b) students who failed on each item (9 in G5A, 15 in G5B, and 21 in G6), or (c) the ones without a distinguishable membership (5 in G5A, 8 in G5B, and 26 in G6).

*5.2.1 Inter-class variation and mathematics performance.* Gender effect on class membership was thus either neglectable or fairly small; gender distribution appeared to be more alike in latent classes of Grade 6 than those of Grade 5. As shown in Table 3, between latent classes, neither G5A ($\chi^2$(1, valid $n$ = 359) = 1.324, $p$ = .250) nor G6 ($\chi^2$(2, valid $n$ = 1150) = 2.283, $p$ = .319) differed

**Table 2.** Mixture Rasch model (MRM) indices for latent class selections.

| Test | No. of classes | AIC | BIC | CAIC |
|---|---|---|---|---|
| G5A | 1 | 3,336 | 3,413 | 3,433 |
| | **2** | 3,197 | **3,357** | **3,398** |
| | 3 | 3,194 | 3,435 | 3,497 |
| G5B | 1 | 11,031 | 11,173 | 11,203 |
| | 2 | 10,444 | 10,733 | 10,794 |
| | **3** | 10,247 | **10,682** | **10,774** |
| | 4 | 10,186 | 10,768 | 10,891 |
| G6 | 1 | 16,627 | 16,779 | 16,809 |
| | 2 | 15,842 | 16,151 | 16,212 |
| | **3** | 15,641 | **16,108** | **16,200** |
| | 4 | 15,508 | 16,132 | 16,255 |

*Note*. BIC = Bayesian Information Criterion. For each test, the number of classes in bold has the smallest BIC value.

**Table 3.** Rational numbers and proportional reasoning performance by latent class.

| Test | Latent class | *N* | % | Probability *Mean* | Gender distribution | | Test results | |
|---|---|---|---|---|---|---|---|---|
| | | | | | $N_{GIRL}$ (%) | $N_{BOY}$ (%) | *Mean* | *SD* |
| G5A | A | 105 | 29 | .84 | 41 (45) | 64 (54) | 450 | 57 |
| | B | 254 | 71 | .94 | 116 (39) | 138 (61) | 431 | 38 |
| | Two classes | 359 | 100 | .89 | 157 (44) | 202 (56) | 436 | 45 |
| G5B | A | 151 | 18 | .84 | 59 (39) | 92 (61) | 500 | 48 |
| | B | 559 | 67 | .92 | 289 (52) | 268 (48) | 449 | 49 |
| | C | 119 | 14 | .91 | 47 (39) | 72 (61) | 441 | 60 |
| | Three classes | 829 | 100 | .89 | 395 (48) | 432 (52) | 457 | 55 |
| G6 | A | 500 | 43 | .85 | 228 (46) | 272 (54) | 552 | 78 |
| | B | 416 | 36 | .79 | 203 (49) | 212 (51) | 515 | 91 |
| | C | 235 | 20 | .85 | 101 (43) | 134 (57) | 496 | 83 |
| | Three classes | 1,151 | 100 | .83 | 532 (46) | 618 (54) | 527 | 87 |

*Note*. Due to missing information on gender, the total number of boys and girls in the latent class C of G5B was not equal to the total number of students affiliated to this latent class. This also applies to the latent class B in the G6 test.

significantly in gender distribution, with significant inter-class variation only identified in G5B ($\chi^2$(2, valid $n$ = 827) = 11.624, $p$ = .003, $\varphi$ = .119).

Significant differences of mathematics performance were found between the latent classes in G5A ($t$ (144) = 3.21, $p$ = .01, $d$ = .41), G5B ($F$(2, 826) = 66.22, $p$ < .001, $\eta_p^2$ = .14), and G6 ($F$(2, 1148) = 42.87, $p$ < .001, $\eta_p^2$ = .07). Post hoc tests of the ANOVA results showed that in G5B, latent class A significantly outperformed latent classes B and C which however did not differ from each other significantly; that all G6 classes performed significantly differently from each other.

*5.2.2 Inter-class variation of item facilities.* The results of chi-square tests of class membership effects on item passing show that, where statistical significance occurred, certain items posed a greater effect

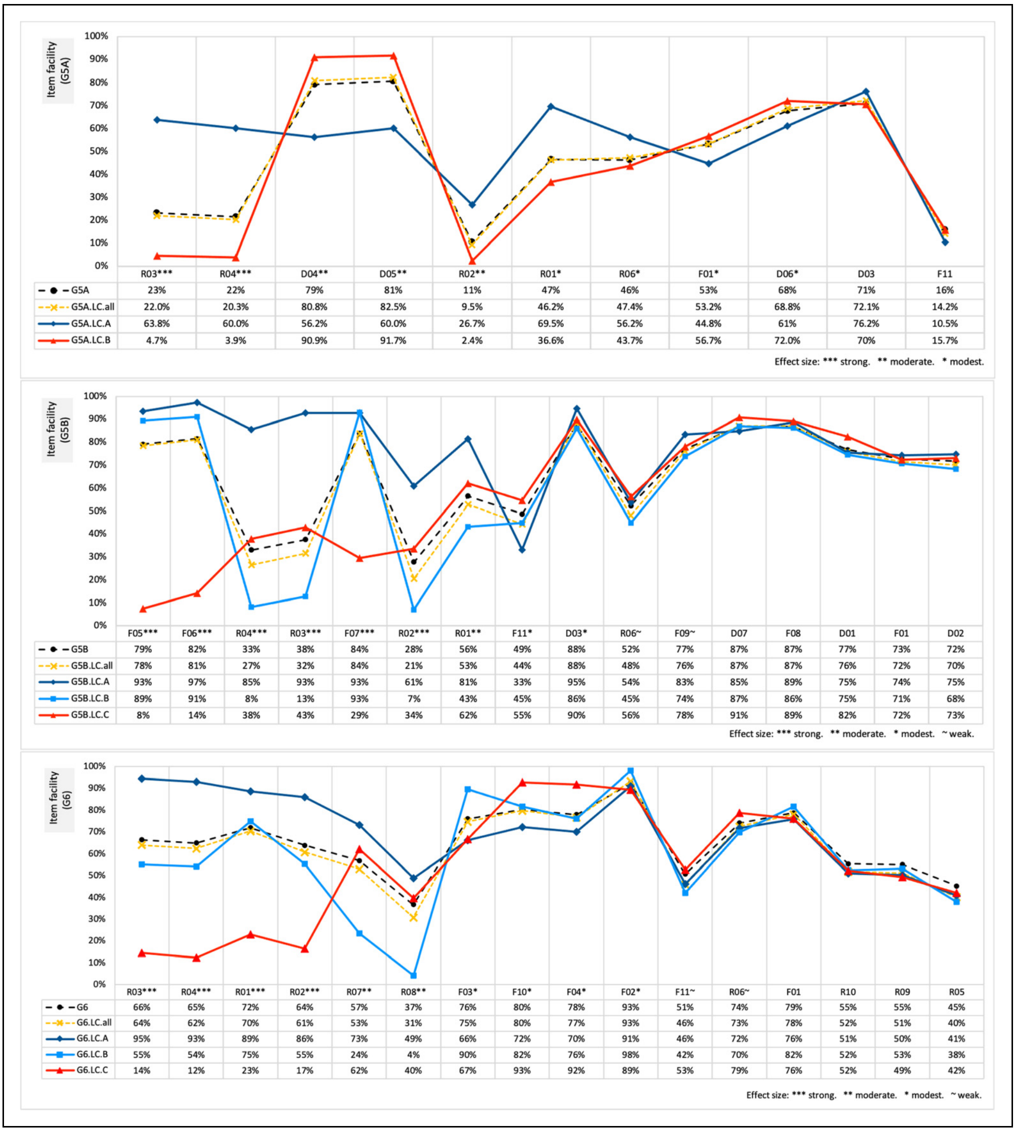



**Figure 5.** Patterns of item facilities across latent classes.

(i.e., φ) than others, and students from various latent classes of G5B and G6 did similar on about a half of the items (Figure 5).

In G5A, the widest inter-class gap lies in two items, R03 and R04 where class A did significantly better (φ = .65 and .63) than class B which, however, outperformed the former on two decimal items, D04 and D05 (φ = .40 and .38). In addition, class A also did considerably better than class B on R01, R02, and R06 (φ = .38 down to .11).

In G5B, in terms of overall mathematics performance, latent class A ($M = 499.6$, $SD = 47.5$) did better than classes B ($M = 449.0$, $SD = 49.4$) and C ($M = 441.0$, $SD = 59.8$) which were, however, not significantly different from each other. At the item level, except F11, class A showed the highest facility on 9 items and a similar facility to classes B and C's on six items. Items F05, F06, R04, R03, F07, and R02 saw the substantial difference between classes, with a strong effect of class membership (φ = .71 to .52); item R01 found a moderate effect (φ = .301), with items F11, D03, R06, and F09 generating a weak effect (φ = .13 to .08) and the remaining four items no difference. In G5B, the widest gap between classes A and C lied in items F05, F06, and F07, with the difference of item facility being 85%, 83%, and 64%, respectively; the widest gap between classes A and B lied in items R02, R03, and R04, with the facility difference being 54%, 80%, and 77%, respectively. The items R03 and R04 also saw a considerable gap (difference in facility: 47% to 50%) between classes A and C. The "eel-feeding" items (R02, R03, and R04) saw substantially wide and almost evenly distributed gaps between classes A and C and between B and C.

In Grade 6, overall, latent class A ($M = 551.9$, $SD = 77.8$) performed significantly better than class B ($M = 514.8$, $SD = 91.4$), and so did class B than C ($M = 495.6$, $SD = 82.7$). At the item level, more than half of 16 items saw significantly different facilities between classes. R03, R04, R02, and R01 (in the descending order in effect sizes) were the items that reflected the biggest inter-class difference with a strong effect (φ = .64 to .54). These were followed by items R07, R08, F03, F10, F04, and F02 where latent class memberships yielded moderate to modest effects (φ = .45 to .15). Items F11 and R06 witnessed a minor difference between latent classes with a rather weak effect (φ = .09 and .08), whereas items F01, R10, R09, and R05 generated quite a common pattern across all latent classes. Item parameters showed similar patterns across latent classes (Figure 6).

In summary, in G5A, class A did better in ratio, whereas class B was better at decimals. In G5B, class A were better at ratio than classes B and C; class B performed the worst in ratio; class C did about the same as classes A and B on all fraction items but F05–F07, the three items that demand students to shade in 2/3 of a shape. However, class C performed better on the eel-feeding items (R01–R04) than class B. Since class C may have lost scores on F05–F07 due to reading inaccuracy (see the corresponding micro diagnosis in the next section), the improvement of accuracy may uplift their overall score and transform the "landscape" of membership distribution in G5B accordingly. In G6, class A had the highest overall score due to their best performance in ratio across classes; class B did better on items R01–R04 than class C but worse on R07 and R08, which presents a mixed picture of ratio competence among them, indicating a need for further improvement about the topic for both classes.

With the picture of children's overall performance and latent class allocations in mind, we now zoom in to look for more detailed patterns of their performance located on the highest end of the mathematical knowledge and thinking hierarchy in rational numbers and proportional reasoning. Our attention focuses specifically on items that received strong effects from the latent class membership in chi-square tests on item passing between classes (i.e., with a φ greater than .5). About these items, microlevel differences were explored in greater depth, which is discussed in the next section. Through the in-depth scrutiny, we sought to diagnose the reasoning processes that had contributed to substantial variation of performance between latent classes.


**Figure 6.** Patterns of item parameters across latent classes.

## *5.3 Micro diagnoses (RQ3): What are the typical reasoning processes generated on the items that strongly distinguish latent classes?*

Both decimal and fraction items had a mean facility of 78% with an *SD* of 7% and 12%, respectively. This partially echoes the fact that no decimal items yielded strong differences between latent classes. Strong differences were however generated by three items in fractions (F05 to F07) and four in ratios and proportions (eel-feeding items: R01 to R04). The seven items were not necessarily the hardest in their topic. The four ratio items were indeed the hardest to fifth graders (facility: 23% to 54%) but not to sixth graders who found R08 (37%) much harder than these (64% to 72%). The fraction items were not the hardest to fifth graders in G5B (79% to 84%) either. A common pattern across the three tests was that items R02–R04 (i.e., the CSMS adaption of Piagetian items) yielded a strong effect on latent class memberships. Significant inter-latent class differences were also strongly yielded by item R01 in G6. In G5B, three fraction items, F05–F07, strongly significantly distinguished latent class C from A and B.

The remaining parts of this section show systematic analyses of multi-category responses that led to in-depth diagnoses of learners' reasoning processes over the course of tackling the seven key items.

*5.3.1 Cognitive diagnoses of key fraction items.* In test G5B, due to misunderstanding of the assessed content, children in latent class C performs significantly differently from classes A and B. For example, items F05 and F06 saw a considerable difference of passing rates between latent classes C (8% of 119) and A (93% of 151) and between C and B (89% of 559). The diagnoses of detailed responses showed two types of major errors.

In latent class C, there were more students having mistakenly shaded in 1/3—rather than the expected 2/3—of the shapes in the two items. The proportions of its members having done so on F05 and F06 (70.6% and 64.7%) was greater than the corresponding proportions for latent classes A (4.6% and 1.3%, $z_{F05} = 11.38$, $p < .001$; $z_{F06} = 11.37$, $p < .001$) and B (5.5% and 3.9%, $z_{F05} = 17.17$, $p < .001$; $z_{F06} = 17.05$, $p < .001$).

Another typical error was shading in 1/2 of the given shapes, which also saw a significant variation between latent classes. Class C had a higher proportion (11.8% on F05 and 10.1% on F06) of members doing so than the proportion of those in classes A (0% and 0.7%; $z_{F05} = 4.33$, $p < .001$; $z_{F06} = 3.73$, $p < .001$) and B (2.1% and 2.0%; $z_{F05} = 4.96$, $p < .001$; $z_{F06} = 4.44$, $p < .001$).

It is unclear whether those shading in 1/3 of the given shape instead of 2/3 were due to a misreading of the expected 2/3 for 1/3. Either way, they were cognitively closer to a correct solution than those shading in 1/2 of the given shape. Knowing how to represent a third is a crucial step toward knowing how to represent two or more thirds.

*5.3.2 Cognitive diagnoses of key ratio items.* In each test, in comparison with other latent classes, the latent class A consisted a much higher proportion of members who took the multiplicative strategy on the eel-feeding items (R01–R04) and carried out the strategy correctly with a complete solution.

Items R01 to R04 form the quartet problem series on "eel feeding" which were originally created by Piaget et al. (1968, 1977) and then adapted in the CSMS "so that the eel lengths were not integer multiples of the smallest" (Hart, 1980, p. 120). The current study adopted the CSMS version, including them as four of the seven common items that all fifth and sixth graders attempted. Amidst those yielding substantial inter-class differences, these were the four that saw a stronger and more evenly spread difference. The abstract proportional problems embedded in R01, R02, R03, and R04 were $10{:}15 = 2{:}\square$ (or $10{:}2 = 15{:}\square$), $15{:}25 = 9{:}\square$ (or $15{:}9 = 25{:}\square$), $10{:}25 = \square{:}10$ (or $25{:}10 = 10{:}\square$), and $15{:}25 = \square{:}10$ (or $25{:}10 = 15{:}\square$), respectively, as modeled in the multiplicative loops in Appendix B.

In the hierarchy of ratios and proportions (four levels) in the CSMS study, R01 was classified as Level 2, and the other three as Level 3. The current study shared a similar finding: R01 was relatively easier than the other three items of this quartet item set. Indeed, as it was intended in the CSMS, the lengths of eels were not integer multiples of the smallest, and such an intention was also realized in the relationship between the lengths of eels and the lengths of fish fingers they should be fed with in R02, R03, and R04 but not in R01 where the length of the shorter eel (10 cm) was 5 times that of the fish finger it was fed with.

Between latent classes, typical errors were comparatively diagnosed through integration of quantitative and qualitative analyses. Quantitative analyses compared percentages of two types of diagnosable errors and overall error rates through $z$ tests (Table 4). Both quantitative and qualitative analyses scrutinized the rationales behind learners' cognitive footsteps (Figures 7 and 8). The combination of quantitative and qualitative analyses helped track children's reasoning processes. About the quartet item set on eel feeding, two major types of reasoning emerged among those who failed on the items: (a) additive (+/–) and (b) multiplicative (×/÷). As shown in the shaded areas in Table 4,

**Table 4.** Diagnoses of errors on the eel-feeding items across latent classes ($z$ tests).

| Item Response | G6 B-A $z$ | $p$ | C-A $z$ | $p$ | B-C $z$ | $p$ | G5B B-A $z$ | $p$ | C-A $z$ | $p$ | B-C $z$ | $p$ | G5A B-A $z$ | $p$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R01 | | | | | | | | | | | | | | |
| +/– [1a] | 1.81 | .07 | 18.93 | *** | 2.25 | * | 1.31 | .19 | 0.42 | .68 | 1.64 | .10 | 3.32 | *** |
| ×/÷ [1b] | 1.19 | .23 | 3.06 | ** | 1.92 | .05 | 1.80 | .07 | 0.79 | .43 | 0.98 | .33 | 1.36 | .17 |
| Error (all) | 5.56 | *** | 17.85 | *** | 12.85 | *** | 8.36 | *** | 3.54 | *** | 3.79 | *** | 5.69 | *** |
| R02 | | | | | | | | | | | | | | |
| +/– [2a] | 4.64 | *** | 4.37 | *** | 0.10 | .92 | 3.56 | *** | 1.28 | .20 | 1.93 | .05 | 3.49 | *** |
| ×/÷ [2b] | 0.13 | .90 | 1.85 | .06 | 1.62 | .10 | 0.48 | .63 | 0.08 | .94 | 0.64 | .53 | 0.53 | .60 |
| Error (all) | 10.31 | *** | 18.26 | *** | 9.63 | *** | 15.17 | *** | 4.46 | *** | 8.22 | *** | 7.15 | *** |
| R03 | | | | | | | | | | | | | | |
| +/– [3a] | - | - | 1.46 | .14 | 1.33 | .18 | 0.18 | .85 | 0.89 | .37 | 0.80 | .42 | 0.64 | .52 |
| ×/÷ [3b] | 3.05 | ** | 3.39 | *** | 0.44 | .66 | 2.23 | * | 1.60 | .11 | 0.90 | .37 | 1.74 | .08 |
| Error (all) | 14.09 | *** | 21.90 | *** | 10.13 | *** | 19.09 | *** | 8.94 | *** | 7.78 | *** | 12.29 | *** |
| R04 | | | | | | | | | | | | | | |
| +/– [4a] | 1.29 | .20 | 2.68 | *** | 1.01 | .31 | 0.84 | .40 | 1.79 | .07 | 1.39 | .16 | 1.95 | .05 |
| ×/÷ [4b] | 2.87 | ** | 1.47 | .14 | 1.19 | .23 | 3.23 | ** | 2.25 | * | 1.07 | .28 | 2.00 | * |
| Error (all) | 13.60 | *** | 25.89 | *** | 10.49 | *** | 19.53 | *** | 8.11 | *** | 8.60 | *** | 12.01 | *** |

*$p$ < .05; **$p$ < .01; ***$p$ < .001.
[1a] (i) $25-10+2=17$, (ii) $10-2=8$, (iii) $25+2=27$; [1b] (i) $2\times 2=4$, (ii) $2\times 3=6$.
[2a] (i) $25-15=10$, (ii) $9+2=11$, (iii) $25-(15-9)=19$; [2b] (i) $9\times 2=18$.
[3a] (i) $10-2=8$; [3b] (i) $10\times 2=20$, (ii) $25\div 10=2.5$, (iii) $2\times 3=6$.
[4a] (i) $10-2=8$, (ii) $10+15=25$; [4b] (i) $10\div 2=5$.

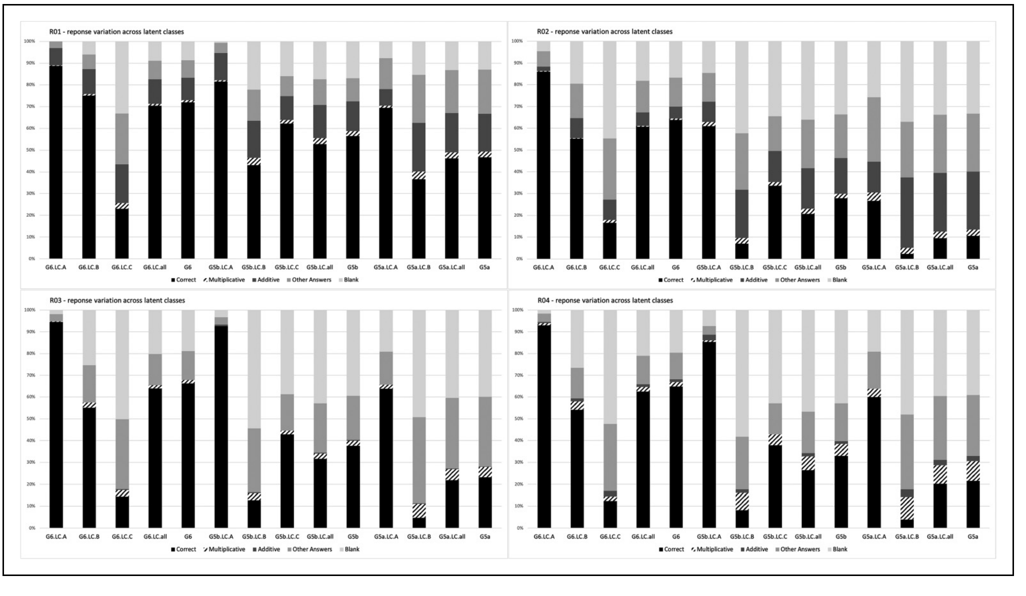



**Figure 7.** Diagnoses of reasoning processes across latent classes on the eel-feeding items.

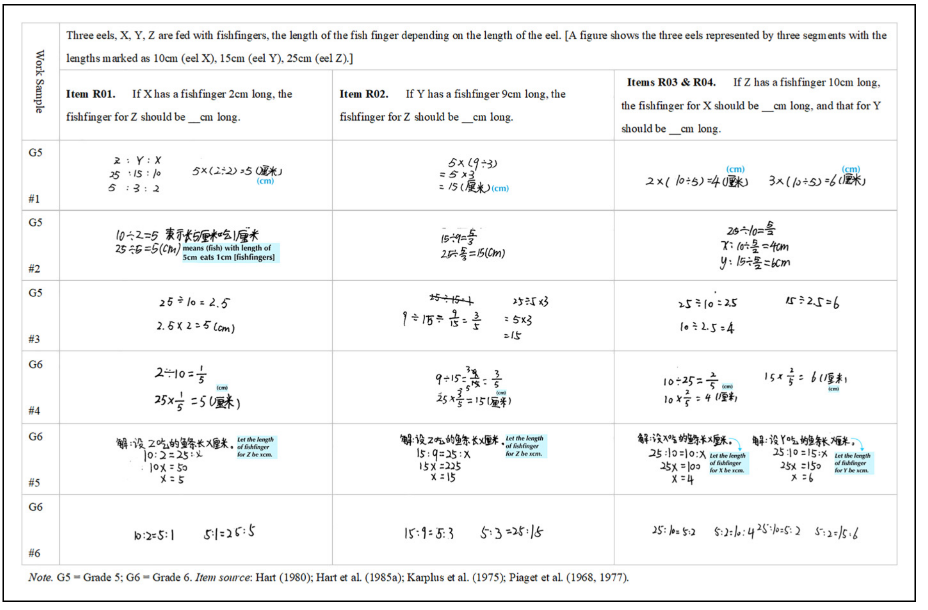



**Figure 8.** Multiple solutions for the four eel-feeding items: Work sample.
*Note.* G5 = Grade 5; G6 = Grade 6.
*Source*: Hart (1980); Hart et al. (1985a); Karplus et al. (1975); Piaget et al. (1968, 1977).

their common rationales, should they have, for modeling the problem the way they did were similar: (1) feeding longer eels with longer fish fingers through adding a certain difference to a certain length or through doubling/tripling and (2) feeding shorter eels with shorter fish fingers by trimming a given length down through subtraction or division.

Among proportionally competent learners in Grades 5 and 6, a rich collection of reasoning processes emerged, falling into the grand patterns captured in Figures 7 and 8. Reasoning about the ratios, fractions, multiplication, or division hidden beneath given and unknown quantities in various ways, fifth graders modeled the problem situation arithmetically, whereas sixth graders did so either arithmetically or algebraically. Ultimately, various approaches took their root in the constant multiplicative loops that connect the quantities (Appendix B). Together, errors and correct solutions shed light on the key toward competent proportional reasoning: being clear about the underlying invariant relationship between varying quantities. With clear identification of the constant yet latent relationship comes fruitful reasoning, in which case it no longer matters whether such constant relationships are phenomenally modeled as ratios, fractions, multiplication or division or a combination of these phenomena.

## 6. Discussion

With relatively hard items of rational numbers and proportional reasoning, the study looked into the performance patterns, latent classes, and reasoning processes of 2,642 students in Chinese Grades 5 and 6 taught by master mathematics teachers. Through the Rasch model, MRM, *z* tests, and other statistical methods, key findings were generated at the macro, meso, and micro-levels, regarding the learning of the topics. In this section, we integrate the key findings with existing research, discuss implications for research and practice, and reflect upon the limitations of the study.

### 6.1 Key findings

At the *macro level*, it was found that children could reach a higher level of mastery in these challenging topics of school mathematics at a younger age, when provided with excellent teaching. With a mean facility of above 3/4, rational number items were on average easier to the learners than were ratio items which had a mean facility of slightly above 1/2. The cross-sectional observation showed a jump of the average facility of ratio items from 38% in Grade 5 to 59% in Grade 6. The current cohort's performance on fractions appears to be better than a recent study with 310 sixth graders from mixed schooling backgrounds in Beijing (Jiang et al., 2021), suggesting differential effects of teaching. Likewise, rational number items, such as the awareness of infinite space between two seemingly adjacent rational numbers, conversion between two adjacent places (similar to equivalent fractions), and fraction division involving unlike denominators, did not appear to be as difficult as were found in the CSMS study (Hart et al., 1981) or the ICCAMS study (Hodgen et al., 2010) where the same items were developed and/or utilized. Overall, the fifth and sixth graders reached a much higher proficiency on each of the 28 items than those in the two studies involving English secondary students 30 years apart, which once again indicates potential effects of teaching on learning.

The cross-sectional data from two adjacent grades show a better performance as well as the disappearance of the gender gap among the senior graders. On the one hand, learners in Grade 6 were indeed more competent than those in Grade 5, with a higher global score and mean facility of items. On the other hand, the commonly found gap in mathematics between girls and boys did exist in the majority of Grade 5 (i.e., G5B), with a weak to moderate effect of gender, but not in Grade 6. This is an encouraging sign for closing the gender gap in senior years with value added by quality teaching.

At the *meso level*, there existed in each test group latent classes manifesting significantly different performance at both the test and specific item levels. Natural number bias did not appear to be an issue as these learners sought to tackle fraction and decimal problems. The greatest difference between latent classes lied in proportional reasoning, as found previously among teachers as learners (Izsák et al., 2010). Albeit based on data sampled differently from previous studies (Hart et al., 1981; Misailidou & Williams, 2003), reasons behind typical errors were similar to what had been found previously—the use of additive or incorrect multiplicative methods. The current study further found it a reasoning feature of the latent class(es) lagging behind others. The systematic analysis of reasoning processes located learners in need of support and diagnosed the particular obstacles that prevented them from mastering the topics.

At the *micro level*, systematic analyses indicated that diagnosable erroneous answers were more likely due to wrong interpretations and modeling of relationships, although calculation errors may still occur with a right model. Cognitive errors tended to be based on addition/subtraction or inaccurate modeling of the relationship with multiplication/division. Fruitful proportional reasoning was more likely to happen as one attempted to seek constant relationships between changing quantities (Hart et al., 1981; Lamon, 2007; Piaget et al., 1977). With a rich perception of relationships between quantities in the multiplicative loops (Appendix B), competent learners in the current study took various approaches to modeling the underlying relationships embedded in the problems, with sixth graders harnessing algebra to tackle these problems, in addition to arithmetic. The cross-sectional trend echoes existing research in that the performance of rational numbers predicts that of algebra (Hurst & Cordes, 2018).

Existing research shows a picture of rich theories and empirical work. However, there are a few "jigsaw pieces" missing from the picture. The study has made contribution to the recovery of these missing pieces.

First, we have studied performance in rational numbers and proportional reasoning among students taught by master mathematics teachers who have been rarely studied at such a scale before.

Drawing on data from children taught by master teachers, the study has gained insight into the extent to which children's competence in these challenging topics of school mathematics can be nurtured when quality teaching is readily available. Systematic research into the nature of mathematics learning in master teachers' classes has shed light on the kind of learning that master teachers can formulate and are likely to transfer into more teachers' classrooms through PD, given their influence among peers in China (Cravens & Wang, 2017; Fan et al., 2015; M. Zhang et al., 2021). Thus, in addition to theoretical contribution, the study may inform mathematics teaching practice and teacher PD with empirical evidence on learning.

Second, the improvement of learning in these topics cannot happen without explicitly locating learning difficulties and rationales behind such difficulties. This makes it important to study both learning outcomes and reasoning processes of learners. With abundant studies looking at the former, relatively few have also looked into the latter (Hart et al., 1981; Piaget et al., 1968; Van Dooren et al., 2010) , but evidence is built more on qualitative data or data collected with secondary schoolers. Seeking quantitative evidence with data from primary schoolers, the current study has pushed the boundaries of research on the topics toward reasoning processes at scale.

Third, the quality of research findings will be more informative to both research and practice if the research design incorporates the latent class perspective. Indeed, latent classes have recently emerged as aims for and means of systematically studying differences in mathematics performance. Yet, such evidence has not been widely accumulated and is limited to either parts of the topics (González-Forte et al., 2021) or teachers as participants (Izsák et al., 2010; Ölmez & Izsák, 2021). More empirical work utilizing latent classes is needed to study the learning of the topics (phenomena) as a connected whole among learners. Under the latent class framework, indications of former research may be further scrutinized. One indication from existing research is that those who failed on the topics are likely to choose additive over multiplicative strategies, which is nevertheless seldom explored within and between latent classes of learners. The dearth of such evidence lies in both quantitative and qualitative terms. The formation of such evidence would make the improvement of learning more focused. Another indication from prior research is that those performing better on the topics have profounder understanding of the connections between fractions, division and ratios than others (Carpenter et al., 1993/2009; Freudenthal, 1983; Lamon, 2007). A latent class design offers opportunity to seek empirical underpinnings for such an indication. The current study makes contribution to the field with respect to the above methodological and theoretical issues.

Last but not least, the variation of learning on these topics needs to be studied more along two important dimensions: gender and grades (years of schooling). Although variation of learning outcomes between gender has been studied globally with mathematics as one subject (e.g., Kaiser & Zhu, 2022; Zhu et al., 2018), it has not been sufficiently covered in research on rational numbers and proportional reasoning. Grades are another dimension of variation that demands more evidence. Such variation should ideally be studied with longitudinal data under an experimental design. In the absence of such an ideal combination, the current study takes the cross-sectional approach, gaining insights into the cognitive development of learners with differential years of schooling (González-Forte et al., 2021).

### *6.2 Implications*

Methodologically, the study goes beyond the otherwise stand-alone statistical methods and integrates them with a clear purpose—to better serve the RQs, proving the practicality and benefit of doing so. For research and practice, findings based on the current study and existing research indicate that profound understanding should be cultivated among learners regarding the connections between fractions, ratios, division and multiplication as the inverse of division (Cai & Wang, 2006; Moseley et al., 2007). Such connectedness allows learners to switch freely between different representations (i.e., $\frac{a}{b}$, $a$: $b$, $a \div b$, or $a \times \frac{1}{b}$, $b \neq 0$) of relationships between quantities according to specific situations (Freudenthal, 1983; Lamon, 2007). To cultivate deeper understanding of the topics among learners, teachers may put more emphasis on the rational number being the relationship between two quantities upon which the sense of the magnitude of any rational number should be built (Torbeyns et al., 2015).

### *6.3 Limitations*

Albeit with rich and robust findings, the study has limitations. It is limited to cross-sectional data with only two grades in primary schools included. The nature of cross-sectional data also means that the study cannot draw absolute causal conclusions which would be better approached with an experimental design. Alternatively, large-scale longitudinal studies following learners from primary through secondary schools could be potential for gaining deeper insight into the mechanism and trend of learner development in this area of school mathematics. Though gender and grade have been taken into account, results in the current study are mainly based on our attention to intrinsic mechanisms in children's cognition that lead to certain reasoning processes and results. Extrinsic factors, such as classroom teaching, are not scrutinized and should be in the future, given recent evidence of positive teaching effects on fraction learning among students with learning difficulties (S. Zhang et al., 2021).

## Acknowledgements

The authors would like to thank all the students and the master teachers for taking part in the project. Our enormous gratitude goes to the following teaching research officials (TROs, in the alphabetic order of provinces): Tao Hu (TRO in Anhui); Cunli Fan, Kechen Liu, and Yanwei Wang (TROs in Beijing); Qingsong Guo and Qihua Zhang (TROs in Jiangsu); Xianqing Song (TRO in Jiangxi); and Zhanjie Ren (TRO in Tianjin) for their kind support in helping us recruit master teachers. We also owe huge gratitude to Yongchun Wang, Chief Editor for Primary Mathematics at the People's Education Press, for introducing TROs in Beijing to us. We are immensely grateful to a number of head teachers and heads of the mathematics department from the participating schools for their kind support during the data collection, though we cannot name them here for confidential concerns. Huge gratitude goes to Professor Jeremy Hodgen at the UCL for permitting us to use the CSMS items. We would also like to thank Jiangxi Normal University for funding the project and the three undergraduate students, Jinying Fu, Yuwen Zhang, and Yuanyuan Zhu for their contribution to the grading of the test. Last but not least, we thank the four anonymous reviewers for their thoughtful and helpful comments with which we were able to further improve the manuscript. The authors take entire responsibility for the quality and findings of the work presented here.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Jiangxi Normal University (grant number 12018578).

## Contributorship

Zhenzhen Miao collected and analyzed the data and drafted the manuscript. Zhenzhen Miao, Christian Bokhove, and David Reynolds designed and implemented the project together. Charalambos Y. Charalambous was invited in the project at a later stage. All authors worked together on the revision and finalization of the manuscript.

## ORCID iDs

Zhenzhen Miao https://orcid.org/0000-0001-8495-7341
Christian Bokhove https://orcid.org/0000-0002-4860-8723

## References

Behr, M. J., Lesh, R., Post, T., & Silver, E. A. (1983). Rational number concepts. In R. Lesh, & M. Landau (Eds.), *Acquisition of mathematics concepts and processes* (pp. 91–126). New York, NY: Academic Press.

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, *35*(5), 463–482. https://doi.org/10.1016/S0883-0355(02)00004-6

Bond, L., Smith, T., Baker, W. K., & Hattie, J. A. (2000). *The certification system of the national board for professional teaching standards: A construct and consequential validity study*. Greensboro, NC: Greensboro Center for Educational Research and Evaluation, University of North Carolina at Greensboro.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. New York, NY: Routledge.

Brown, M., Küchemann, D. E., & Hodgen, J. (2010). The struggle to achieve multiplicative reasoning 11-14. In M. Joubert, & P. Andrews (Eds.), *Proceedings of the Seventh British Congress of Mathematics Education (BCME7)* (Vol. 30, pp. 49–56). Manchester, UK: BSRLM.

Cai, J., & Wang, T. (2006). U.S. and Chinese teachers' conceptions and constructions of representations: A case of teaching ratio concept. *International Journal of Science and Mathematics Education*, *4*(1), 145–186. https://doi.org/10.1007/s10763-005-9006-7

Carpenter, T. P., Fennema, E., & Romberg, T. A. (Eds.) (2009). *Rational numbers: An integration of research*. New York, NY: Routledge. (Original work published 1993)

Charalambous, C. Y. (2016). Investigating the knowledge needed for teaching mathematics: An exploratory validation study focusing on teaching practices. *Journal of Teacher Education*, *67*(3), 220–237. https://doi.org/10.1177/0022487116634168

Charalambous, C. Y., & Pitta-Pantazi, D. (2005). Revisiting a theoretical model on fractions: Implications for teaching and research. In H. L. Chick, & J. L. Vincent (Eds.), *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 233–240). Melbourne, Australia: PME29.

Charalambous, C. Y., & Pitta-Pantazi, D. (2007). Drawing on a theoretical model to study students' understandings of fractions. *Educational Studies in Mathematics*, *64*(3), 293–316. https://doi.org/10.1007/s10649-006-9036-2

Copur-Gencturk, Y. (2021). Teachers' knowledge of fraction magnitude. *International Journal of Science and Mathematics Education*, *20*(5), 1021–1036. https://doi.org/10.1007/s10763-021-10173-2

Cravens, X., & Wang, J. (2017). Learning from the masters: Shanghai's teacher-expertise infusion system. *International Journal for Lesson and Learning Studies*, *6*(4), 306–320. https://doi.org/10.1108/IJLLS-12-2016-0061

Empson, S. B., Levi, L., & Carpenter, T. P. (2011). The algebraic nature of fractions: Developing relational thinking in elementary school. In J. Cai, & E. Knuth (Eds.), *Early algebraization: A global dialogue from multiple perspectives* (pp. 409–428). Heidelberg, Germany: Springer-Verlag Berlin Heidelberg.

Fan, L., Zhu, Y., & Tang, C. (2015). What makes a master teacher? A study of thirty-one mathematics master teachers in Chinese mainland. In L. Fan, N.-Y. Wong, J. Cai, & S. Li (Eds.), *How Chinese teach mathematics* (pp. 493–528). Singapore: World Scientific.

Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht, The Netherlands: D. Reidel Publishing.

Freudenthal, H. (1978). *Weeding and sowing: Preface to a science of mathematics education*. Dordrecht, The Netherlands: D. Reidel Publishing.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht, The Netherlands: D. Reidel Publishing.

González-Forte, J. M., Fernández, C., Van Hoof, J., & Van Dooren, W. (2021). Profiles in understanding operations with rational numbers. *Mathematical Thinking and Learning*, *24*(3), 230–247. https://doi.org/10.1080/10986065.2021.1882287

Hart, K. (1980). *Secondary school-children's understanding of ratio and proportion* [Doctoral dissertation, University of London]. British Library EthOS.

Hart, K. (1987). Strategies and errors in secondary mathematics. *Mathematics in School*, *16*(2), 14–17. https://www.jstor.org/stable/30214189

Hart, K., Brown, M., Kerslake, D., Küchemann, D. E., & Ruddock, G. (1985a). *Chelsea diagnostic mathematics tests*. Windsor, UK: NFER-Nelson.

Hart, K., Brown, M., Kerslake, D., Küchemann, D. E., & Ruddock, G. (1985b). *Chelsea diagnostic mathematics tests: Teacher's guide*. Windsor, UK: NFER-Nelson.

Hart, K., Brown, M., Küchemann, D. E., Kerslake, D., Ruddock, G., & McCartney, M. (1981). *Children's understanding of mathematics: 11-16*. London, UK: John Murray.

Hiebert, J., & Behr, M. (1988). *Number concepts and operations in the middle grades*. Reston, VA: National Council of Teachers of Mathematics.

Hodgen, J., Küchemann, D., Brown, M., & Coe, R. (2010). Multiplicative reasoning, ratio and decimals: A 30-year comparison of lower secondary students' understandings. In M. M. F. Pinto, & T. F. Kawasaki (Eds.), *Proceedings of the 34th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 89–96). Belo Horizonte, Brazil: PME.

Howe, C., Luthman, S., Ruthven, K., Mercer, N., Hofmann, R., Ilie, S., & Guardia, P. (2015). Rational number and proportional reasoning in early secondary school: Towards principled improvement in mathematics. *Research in Mathematics Education*, *17*(1), 38–56. https://doi.org/10.1080/14794802.2015.1019914

Huang, R., Ye, L., & Prince, K. (2017). Professional development of secondary matehamtics teachers in mainland China. In B. Kaur, O. N. Kwon, & Y. H. Leong (Eds.), *Professional development of mathematics teachers: An Asian perspective* (pp. 17–31). Singapore: Springer Nature Singapore.

Hurst, M., & Cordes, S. (2018). A systematic investigation of the link between rational number processing and algebra ability. *British Journal of Psychology*, *109*(1), 99–117. https://doi.org/10.1111/bjop.12244

Izsák, A., Orrill, C. H., Cohen, A. S., & Brown, R. E. (2010). Measuring middle grades teachers' understanding of rational numbers with the mixture Rasch model. *The Elementary School Journal*, *110*(3), 279–300. https://doi.org/10.1086/648979

Jiang, Z., Mok, I. A. C., & Li, J. (2021). Chinese students' hierarchical understanding of part-whole and measure subconstructs. *International Journal of Science and Mathematics Education*, *19*(7), 1441–1461. https://doi.org/10.1007/s10763-020-10118-1

Kaiser, G., & Zhu, Y. (2022). Gender differences in mathematics achievement: A secondary analysis of programme for international student assessment data from Shanghai. *Asian Journal for Mathematics Education*, *1*(1), 115–130. https://doi.org/10.1177/27527263221091373

Karplus, R., Karplus, E., Formisano, M., & Paulsen, A.-C. (1975). *Proportional reasoning and control of variables in seven countries*. Berkeley, CA: Lawrence Hall of Science, University of California.

Kiefer, T., Robitzsch, A., & Wu, M. (2021). *TAM (Test analysis modules): An R package*. TAM Tutorials. http://www.edmeasurementsurveys.com/TAM/Tutorials/

Lamon, S. J. (2007). Rational numbers and proportional reasoning: Toward a theoretical framework for research. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 629–667). Reston, VA: National Council of Teachers of Mathematics.

Leinhardt, G. (1989). Math lessons: A contrast of novice and expert competence. *Journal for Research in Mathematics Education*, *20*(1), 52–75. https://doi.org/10.2307/749098

Li, F., Cohen, A. S., Kim, S.-H., & Cho, S.-J. (2009). Model selection methods for mixture dichotomous IRT models. *Applied Psychological Measurement*, *33*(5), 353–373. https://doi.org/10.1177/0146621608326422

Li, Y., Chen, X., & An, S. (2009). Conceptualizing and organizing content for teaching and learning in selected Chinese, Japanese and US mathematics textbooks: The case of fraction division. *ZDM–The International Journal on Mathematics Education*, *41*(6), 809–826. https://doi.org/10.1007/s11858-009-0177-5

Li, Y., Tang, C., & Gong, Z. (2011). Improving teacher expertise through master teacher work stations: A case study. *ZDM–The International Journal on Mathematics Education*, *43*(6), 763–776. https://doi.org/10.1007/s11858-011-0344-3

Liu, R.-D., Ding, Y., Zong, M., & Zhang, D. (2014). Concept development of decimals in Chinese elementary students: A conceptual change approach. *School Science and Mathematics*, *114*(7), 326–338. https://doi.org/10.1111/ssm.12085

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum.

Ministry of Education China (2011). *Mathematics curriculum standards for nine-year compulsory education*. Beijing, China: Beijing Normal University. [in Chinese].

Misailidou, C., & Williams, J. (2003). Diagnostic assessment of children's proportional reasoning. *The Journal of Mathematical Behavior*, *22*(3), 335–368. https://doi.org/10.1016/S0732-3123(03)00025-7

Moseley, B. J., Okamoto, Y., & Ishida, J. (2007). Comparing US and Japanese elementary school teachers' facility for linking rational number representations. *International Journal of Science and Mathematics Education*, *5*(1), 165–185. https://doi.org/10.1007/s10763-006-9040-0

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

National Bureau of Statistics of China (2019). *Gross domestic product of major cities in 2018*. https://data.stats.gov.cn/english/easyquery.htm?cn=E0105.

National Mathematics Advisory Panel (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

Nunes, T., Bryant, P., & Watson, A. (2009). *Key understandings in mathematics learning: Understanding rational numbers and intensive quantities*. London, UK: Nuffield Foundation.

OECD (2019). *PISA 2018 results: What students know and can do (Vol. 1)*. Paris, France: The OECD Publishing.

Ölmez, İ. B, & Izsák, A. (2021). Characterizing reasoning about fraction arithmetic of middle grades teachers in three latent classes. *Mathematical Thinking and Learning*, *23*(3), 225–253. https://doi.org/10.1080/10986065.2020.1780368

Piaget, J., Grize, J. B., Szeminska, A., & Bang, V. (1968). *Epistémologie et psychologie de la fonction*. Paris, France: Presses Universitaires de France.

Piaget, J., Grize, J. B., Szeminska, A., & Bang, V. (1977). *Epistemology and psychology of functions*. Dordrecht, The Netherlands: D. Reidel Publishing.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, *14*(3), 271–282. https://doi.org/10.1177/014662169001400305

Seol, H. (2020). *snowRMM: Rasch Mixture Model for jamovi*. [jamovi module]. GitHub. https://github.com/hyunsooseol/snowRMM

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., & Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science*, *23*(7), 691–697. https://doi.org/10.1177/0956797612440101

The jamovi project (2021). *jamovi*. (Version 2.0) [Computer software]. jamovi. https://www.jamovi.org

Thompson, P. W., & Thompson, A. G. (1994). Talking about rates conceptually, Part I: A teacher's struggle. *Journal for Research in Mathematics Education*, *25*(3), 279–303. https://doi.org/10.2307/749339

Torbeyns, J., Schneider, M., Xin, Z., & Siegler, R. S. (2015). Bridging the gap: Fraction understanding is central to mathematics achievement in students from three different continents. *Learning and Instruction*, *37*, 5–13. https://doi.org/10.1016/j.learninstruc.2014.03.002

Valverde, G. A., Bianchi, L. J., Wolfe, R. G., Schmidt, W. H., & Houang, R. T. (2002). *According to the book: Using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Van Dooren, W., De Bock, D., & Verschaffel, L. (2010). From addition to multiplication … and back: The development of students' additive and multiplicative reasoning skills. *Cognition and Instruction*, *28*(3), 360–381. https://doi.org/10.1080/07370008.2010.488306

Wu, M., Tam, H. P., & Jen, T.-H. (2016). *Educational measurement for applied researchers: Theory into practice*. Singapore: Springer Nature Singapore.

Yang, D.-C., Reys, R. E., & Wu, L.-L. (2010). Comparing the development of fractions in the fifth- and sixth-graders' textbooks of Singapore, Taiwan, and the USA. *School Science and Mathematics*, *110*(3), 118–127. https://doi.org/10.1111/j.1949-8594.2010.00015.x

Zhang, M., Walker, A. D., & Qian, H. (2021). Master teachers as system leaders of professional learning: Master teacher studios in China. *International Journal of Educational Management*, *35*(7), 1333–1346. https://doi.org/10.1108/IJEM-04-2021-0127

Zhang, S., Yu, S., Xiao, J., Liu, Y., & Jiang, T. (2021). The effects of concrete-representational-abstract sequence instruction on fractions for Chinese elementary students with mathematics learning disabilities. *International Journal of Science and Mathematics Education*, *20*(5), 1481–1498. https://doi.org/10.1007/s10763-021-10215-9

Zhu, Y., Kaiser, G., & Cai, J. (2018). Gender equity in mathematical achievement: The case of China. *Educational Studies in Mathematics*, *99*(3), 245–260. https://doi.org/10.1007/s10649-018-9846-z

## Author biographies

**Zhenzhen Miao** is an Assistant Professor in Mathematics Education at Jiangxi Normal University, China. She did her PhD (2016) on the effectiveness of mathematics teaching in primary schools in China and England at the University of Southampton, UK. Upon completing her PhD, she returned to China and took up an Assistant Professor position at Jiangxi Normal University in 2017, under the University's Outstanding Returnee Scholar scheme. The current project was funded by a grant (CNY 100,000) that came with the scheme. Zhenzhen teaches both initial teacher education (ITE) and research courses in Mathematics Education to both undergraduates and postgraduates and is an active contributor to teacher professional development (PD) at local schools, provincial teaching research conferences, and the National PD Programme for Rural Teachers in China. Her research focuses on mathematics learning and teaching as well as mathematics ITE and teacher PD in the local, national, and international contexts.

**Christian Bokhove** is a Professor in Mathematics Education at the Southampton Education School, University of Southampton, UK. He specializes in secondary mathematics education, with an emphasis on international comparisons, as well as using innovative research methods to answer relevant educational research questions.

**David Reynolds (CBE)** is a Distinguished Professor at School of Education, Hangzhou Normal University, China. He has published over 20 books and more than 200 articles and book chapters. He is one of the founders of the field of School Effectiveness and School Improvement (SESI), the SESI Journal, and the International Congress for School Effectiveness and Improvement (ICSEI). David has advised, lectured, and/or consulted in over 60 countries over the years.

**Charalambos Y. Charalambous** is an Associate Professor in Educational Research and Evaluation at the Department of Education of the University of Cyprus. His research interests include teaching effectiveness with a particular focus on understanding the work of teaching, measuring teaching quality, and exploring its relation to student learning.

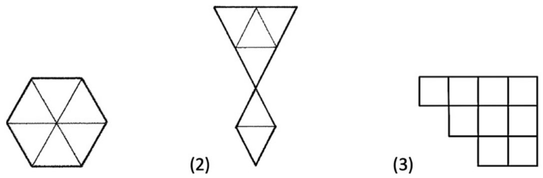
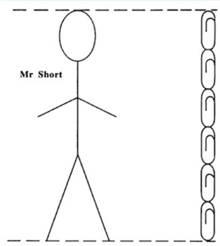
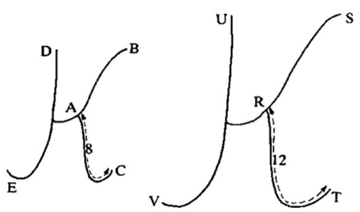
# Appendix A

**Table A1.** An overview of 28 items utilized in the study with item examples.

| Item ID | Item ID in CSMS | Item description (simplified) | CSMS level | Level criteria | Booklet inclusion |
|---|---|---|---|---|---|
| D01 | D11c | Write 11 tenths in decimals. | 5 of 6 | 4/6 | G5B |
| D02 | D11d | How many hundredths are there in four tenths? | 5 of 6 | 4/6 | G5B |
| D03 | D12e | How many numbers are there between 0.41 and 0.42? | 6 of 6 | 3/4 | G5A; G5B |
| D04 | D14g | 24 ÷ 20 | 6 of 6 | 3/4 | G5A |
| D05 | D14h | 16 ÷ 20 | 6 of 6 | 3/4 | G5A |
| D06 | D18f | 59 ÷ 190 is approximately __. | 6 of 6 | 3/4 | G5A |
| D07 | D1b | The number 2 in 521400 represents __. | 5 of 6 | 4/6 | G5B |
| F01 | F1_20<br>F2_15 | Fraction of the floor tiled (9 of 24 equal parts). | F1: 4of4<br>F2: 3of4 | F1: 3/4<br>F2: 2/3 | G5A; G5B; G6 |
| F02 | F1_21a | $\frac{3}{5}$ of wages paid in tax. ⇒ __of wages left. | F1: 2of4 | 7/10 | G6 |
| F03 | F1_21b | $\frac{3}{5}$ of wages paid in tax and $\frac{1}{10}$ on rent. ⇒ __of wages left. | F1: 4of4 | 3/4 | G6 |
| F04 | F1_23 | Number of $1\frac{1}{4}$ cm long sticks that a $8\frac{3}{4}$ cm long stick can be cut into. | F1: 4of4 | 3/4 | G6 |
| F05 | F1_4b | Shade $\frac{2}{3}$ of a hexagon (of 6 equal parts). | F1: 3of4 | 7/11 | G5B |
| F06 | F1_4c | Shade $\frac{2}{3}$ of an 8-like shape (of 6 equal parts). | F1: 3of4 | 7/11 | G5B |
| F07 | F1_4d | Shade $\frac{2}{3}$ of an irregular shape (of 9 equal parts). | F1: 3of4 | 7/11 | G5B |
| F08 | F1_16a | Equivalent fraction: $\frac{2}{7}=\frac{\square}{14}$. | F1: 3of4 | 7/11 | G5B |
| F09 | F1_16b<br>F2_10b | Equivalent fraction: $\frac{2}{7}=\frac{10}{\Delta}$. | F1: 4 of 4<br>F2: 3 of 4 | F1: 3/4<br>F2: 2/3 | G5B |
| F10 | F2_22 | The length of a rectangle, given area $\frac{1}{3}$ and width $\frac{3}{5}$. | F2: 4of4 | 3/5 | G6 |
| F11 | F2_27 | Given $\frac{a}{b} < \frac{a}{c}$, and a, b, and c are positive whole numbers. Write the condition. | F2: 4of4 | 3/5 | G5A; G5B; G6 |
| R01 | R2d | The 10 cm eel eats 2 cm fish fingers. ⇒ The 25 cm eel eats __ cm fish fingers. | 2 of 4 | 3/5 | G5A; G5B; G6 |
| R02 | R2e | The 15 cm eel eats 9 cm fish fingers. ⇒ The 25 cm eel eats __ cm fish fingers. | 3 of 4 | 4/6 | G5A; G5B; G6 |
| R03 | R2f1 | The 25 cm eel eats 10 cm fish fingers. ⇒ The 10 cm eel eats __ cm fish fingers. | 3 of 4 | 4/6 | G5A; G5B; G6 |
| R04 | R2f2 | The 25 cm eel eats 10 cm fish fingers. ⇒ The 15 cm eel eats __ cm fish fingers. | 3 of 4 | 4/6 | G5A; G5B; G6 |
| R05 | R4b | The length of a missing vertical line segment (2 : 3 = __ : 5) | 4 of 4 | 3/4 | G6 |
| R06 | R5 | Mr Short's height = 6 paperclips (4 toothpicks). ⇒ Mr Tall's height = __ paperclips (6 toothpicks). | 3 of 4 | 4/6 | G5A; G5B; G6 |
| R07 | R6a | 1 (honey) : 5 (apple J), 3 (pear J) : 10 (apple J), and 8 (peach J) : 15 (apple J). ⇒__(honey) : __(pear J). | N/A | N/A | G6 |
| R08 | R6b | 1 (honey) : 5 (apple J), 3 (pear J) : 10 (apple J), and 8 (peach J) : 15 (apple J). ⇒__(peach J) : __(pear J). | 4 of 4 | 3/4 | G6 |
| R09 | R7a | Length of RS in the big K (8 : 12 = 9 : __). | 4 of 4 | 3/4 | G6 |
| R10 | R7b | Length of DE in the small K (8 : 12 = __ : 18). | 4 of 4 | 3/4 | G6 |

*Note*. CSMS = Concept in Secondary Mathematics and Science; D = decimals; F = fractions; R = ratios and proportions; G5A = test for fifth graders in Anhui province; G5B = test for fifth graders in four other provinces; G6 = sixth graders in all five provinces, that is, Anhui, Beijing, Jiangsu, Jiangxi, and Tianjin; J = Juice.
*Source*: Hart et al. (1985a, 1985b); with reuse permission from the ICCAMS project PI.

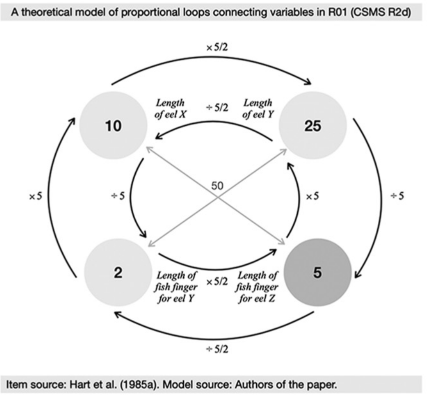
*Item Examples*

| | |
|---|---|
| **F02**<br>**F03**<br><br>CSMS<br>F1-21a<br>F1-21b | Zhang Jie pays $\frac{3}{5}$ of his wages in tax. What fraction of his wages does he have left? __________<br>**Explain your solution:** (SPACE LEFT BLANK FOR WRITING)<br>He also pays $\frac{1}{10}$ of his wages on rent. What fraction of his wages does he have left after tax and rent have been paid? __________<br>**Explain your solution:** (SPACE LEFT BLANK FOR WRITING) |
| **F04**<br>CSMS<br>F1-23 | How many pieces of wood $1\frac{1}{4}$ cm long can we get from a piece of $8\frac{3}{4}$ cm long? __________<br>**Explain your solution:** (SPACE LEFT BLANK FOR WRITING) |
| **F05**<br>**F06**<br>**F07**<br><br>CSMS<br>F1-4b<br>F1-4c<br>F1-4d | Shade in two-thirds of each of these shapes.<br><br> |
| **R06**<br><br>CSMS<br>R5 | You can see the height of Mr Short measured with paper-clips.<br><br>Mr Short has a friend Mr Tall.<br>When we measure their heights with toothpicks:<br>Mr Short's height is four toothpicks.<br>Mr Tall's height is six toothpicks.<br>How many paper-clips are needed for Mr Tall's height?<br>__________<br>**Please explain your solution in the space below and then write your answer on the line above.**<br>(SPACE LEFT BLANK FOR WRITING)<br><br> |
| **R07**<br>**R08**<br><br>CSMS<br>R6ab | To make a particular type of juice drink, the following ingredients are needed:<br>1 spoon of honey to 5 spoons of apple juice,<br>3 spoons of pear juice to 10 spoons of apple juice,<br>8 spoons of peach juice to 15 spoons of apple juice.<br>(1) In the drink, the proportion of honey to pear juice is:<br>__________ spoons of honey to __________ spoons of pear juice.<br>**Explain your solution:** (SPACE LEFT BLANK FOR WRITING)<br>(2) In the drink, the proportion of peach juice to pear juice is:<br>__________ spoons of peach juice to __________ spoons of pear juice.<br>**Explain your solution:** (SPACE LEFT BLANK FOR WRITING) |
| **R09**<br>**R10**<br><br>CSMS<br>R7ab | These 2 letters are the same shape, one is larger than the other. The curve AC is 8 units. RT is 12 units.<br>(1) The curve AB is 9 units. How long is the curve RS? _______<br>**Explain your solution:**<br>(SPACE LEFT BLANK FOR WRITING)<br>(2) The curve UV is 18 units. How long is the curve DE? _______<br>**Explain your solution:**<br>(SPACE LEFT BLANK FOR WRITING)<br><br> |

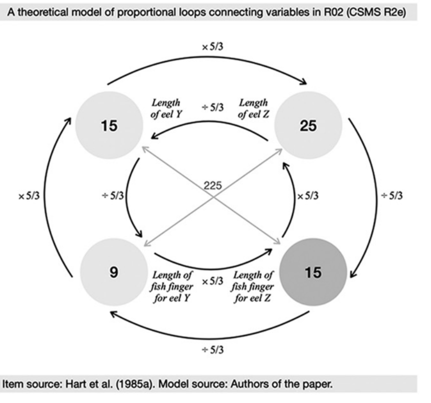
*Source*: Hart et al. (1985a, 1985b); with reuse permission from the ICCAMS project PI.

# Appendix B.

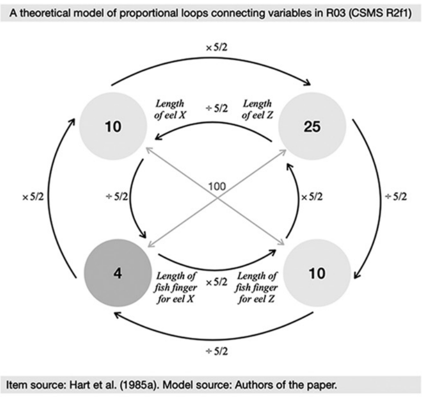
a.



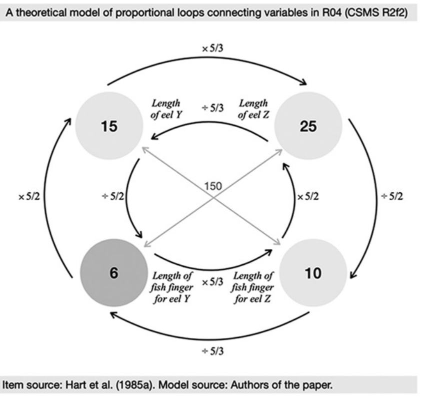

b.




c.




d.




Constant multiplicative loops in the four eel-feeding items.