PROPORTIONAL REASONING PROBLEMS:
CURRENT STATE AND A POSSIBLE FUTURE DIRECTION
Hartono Tjoe Jimmy de la Torre
Rutgers University Rutgers University
hartono.tjoe@rutgers.edu j.delatorre@rutgers.edu

This article surveys the types of proportional reasoning (PR) problems in existing mathematics textbooks, statewide examinations, and international assessments on the subject of mathematics at the middle school level. It examines these PR problems from a newly emerging psychometric framework—cognitive diagnosis models (CDMs)—and suggests a different way in which PR assessments can benefit from such a framework. It will be argued that a direct application of CDMs to the current state of PR problems can be suboptimal, if not counterproductive. It concludes with two cognitively-based PR items to exemplify an alternative direction in the assessment and testing of PR using CDMs.

Keywords: proportional reasoning, ratio, proportion, cognitive diagnosis models

INTRODUCTION

Proportional reasoning (PR) is often viewed as the “capstone of elementary arithmetic, number, and measurement concepts” and the “cornerstone of algebra and other higher level areas of mathematics” (Lesh, Post, & Behr, 1988, p. 97). The National Council of Teachers of Mathematics (NCTM) maintains that a significant amount of middle school mathematics problem-solving should be contextualized towards students’ development of number sense and computational fluency in PR. The concept of PR is “an important integrative thread that connects many of the mathematics topics studied in grades 6-8” (NCTM, 2000, p. 217). Because of its significant role in predicting the degree to which middle school students will succeed in mathematics learning at the high school level, PR naturally becomes one of the main foci in middle school mathematics assessments.

The purpose of this article is to survey the types of PR problems in existing mathematics textbooks, statewide examinations, and international assessments on the subject of mathematics at the middle school level (grade 8, ages 13-14). As will be shown, most existing PR problems measure only a very narrow aspect of how students reason proportionally. Given this state, it will be argued that a direct application of cognitive diagnosis models (CDMs)—a newly emerging psychometric framework—to existing PR problems can be suboptimal, if not counterproductive. This article concludes with two exemplars of cognitively-based PR items, which were constructed and validated using the psychometric characteristics identified to measure specifically how students reason proportionally.
CURRENT STATE OF PROPORTIONAL REASONING PROBLEMS

Most grade 6-8 mathematics textbooks contain the teaching of PR in a chapter labeled “Ratio, Proportion, and Percent,” while international assessments, such as the Trends in International Mathematics and Science Study (TIMSS), evaluate PR under “Fractions and Proportionality.” Although the study of PR is commonly associated with rational numbers, this article places greater emphasis on examining PR problems, and not as much on fractions-and-decimals problems exclusively (e.g., fraction subtraction problems). In addition, as indicated by comprehensive reviews on PR such as the one by Tourniaire and Pulos (1985) and more recently by de la Torre et al. (under review), this article utilizes three main classifications of PR problems: a) missing value problems (MVPs), b) ratio-proportion construction problems, and c) ratios comparison problems.

The researchers examined the following sources of PR problems: a) the “Ratio, Proportion, and Percent” chapters of three grades 6-8 mathematics textbooks that are most commonly used in northeastern public school systems (Larson et al., 2004); b) two sets of state mathematics tests: one year of the New Jersey Assessment of Skills and Knowledge sample test form (NJSDOE, 2001) and six years of New York State grade 8 Regents exams (NYSED, 2005-2010); and c) two sets of international assessments: one year of Programme for International Student Assessment sample questions (PISA, 2009) and four years of TIMSS 1995 and 2003 grade 4 and 1999 and 2003 grade 8 released items (TIMSS, 2010). Examples of the three common types of PR problems are shown in Table 1.

<table>
<thead>
<tr>
<th>Item 1: Missing Value Problem (Item M032533, page 205, grade 8, 2003; TIMSS, 2010)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A machine uses 2.4 liters of gasoline for every 30 hours of operation. How many liters of gasoline will the machine use in 100 hours?</td>
</tr>
<tr>
<td>A. 7.2</td>
</tr>
<tr>
<td>B. 8.0</td>
</tr>
<tr>
<td>C. 8.4</td>
</tr>
<tr>
<td>D. 9.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item 2: Ratio-Proportion Construction Problem (Item 21, page 435; Larson et al., 2004)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your face uses about 13 muscles to smile and about 43 muscles to frown. Write the ratio of muscles used for smiling to muscles used for frowning.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item 3: Ratios Comparison Problem (Item 39, page 7, course 3; NYSED, 2010)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ms. Lembright ordered the types of muffins below for a class party.</td>
</tr>
<tr>
<td>- 12 blueberry muffins for $7.20</td>
</tr>
<tr>
<td>- 8 chocolate muffins for $4.40</td>
</tr>
<tr>
<td>- 6 raisin muffins for $4.50</td>
</tr>
<tr>
<td>What type of muffin costs the least?</td>
</tr>
</tbody>
</table>
From the seven sources, 1,612 PR problems were reviewed and classified into one of the three types of PR problems. The findings revealed that on average, at least 50% of PR problems taught in a classroom setting or tested at a national or international assessment were classified as MVPs. Table 2 presents a summary of the current state of PR problems.

Table 2: Summary of the current state of PR problems

<table>
<thead>
<tr>
<th>Source</th>
<th>Missing Value Problems</th>
<th>Ratio-Proportion Construction Problems</th>
<th>Ratios Comparison Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Textbook Grade 6</td>
<td>202 (50%)</td>
<td>160 (39%)</td>
<td>45 (11%)</td>
</tr>
<tr>
<td>Textbook Grade 7</td>
<td>480 (73%)</td>
<td>128 (19%)</td>
<td>55 (8%)</td>
</tr>
<tr>
<td>Textbook Grade 8</td>
<td>329 (81%)</td>
<td>56 (14%)</td>
<td>20 (5%)</td>
</tr>
<tr>
<td>New Jersey Test</td>
<td>5 (50%)</td>
<td>2 (20%)</td>
<td>3 (30%)</td>
</tr>
<tr>
<td>New York Test</td>
<td>51 (87%)</td>
<td>3 (5%)</td>
<td>5 (8%)</td>
</tr>
<tr>
<td>PISA</td>
<td>17 (94%)</td>
<td>0 (0%)</td>
<td>1 (6%)</td>
</tr>
<tr>
<td>TIMSS</td>
<td>31 (62%)</td>
<td>8 (16%)</td>
<td>11 (22%)</td>
</tr>
<tr>
<td>Total</td>
<td>1,115 (69%)</td>
<td>357 (22%)</td>
<td>140 (9%)</td>
</tr>
</tbody>
</table>

MOVING AWAY FROM SUMMATIVE ASSESSMENTS

Despite their efforts to conduct a comprehensive assessment on students’ mathematics performance, mathematics educators, along with psychometricians for that matter, appear to have lagged behind cognitive psychologists in their progress to understand students’ underlying competence that can be transferred into further learning experiences. Mislevy (1993) states, “It is only a slight exaggeration to describe the test theory that dominates educational measurement today as the application of 20th century statistics to 19th century psychology” (p. 19). Indeed, present large-scale assessments in mathematics are generally grounded on traditional psychometric models.

Traditional psychometric models, such as the classical test theory or unidimensional item response theory (IRT), assign to each student a raw or scaled score (Pellegrino, Chudowsky, & Glaser, 2001). All items tested through those assessment methods place students relative to each other on one continuous latent construct, for example, students’ mathematical proficiencies. Because every student is associated with a single test score, one can conveniently perform a summative assessment.

At a microscopic level, however, although IRT ascribes to an equal probability of answering an item correctly to students with similar mathematical proficiencies, the difficulty level of an item may be perceived differently by different students, depending on the construction of the item itself. In this manner, focusing on superficial features of an item rather than on the item
in its entirety can lead to a highly misleading inference about students’ actual abilities. The
square-root-approximation problem presented in Table 3 provides an illustration of a
multiple-choice item with a single stem and two sets of options of different difficulty levels.
Clearly, a student with an average mathematical proficiency can answer the item with Set A
just as accurately as a student with a high mathematical proficiency.

Table 3: A square-root-approximation problem

<table>
<thead>
<tr>
<th></th>
<th>Set A</th>
<th>Set B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. 4.5</td>
<td>A. 5.4</td>
<td></td>
</tr>
<tr>
<td>B. 5.0</td>
<td>B. 5.5</td>
<td></td>
</tr>
<tr>
<td>C. 5.5</td>
<td>C. 5.6</td>
<td></td>
</tr>
<tr>
<td>D. 6.0</td>
<td>D. 5.7</td>
<td></td>
</tr>
</tbody>
</table>

Still, a test based on the current state of PR problems, as discussed in the earlier section, can
readily provide information about how well a student performs on that test. Aside from the
potentially narrower construct, little is known about how to use such scores to improve
students’ PR skills. For example, if a student cannot answer an MVP correctly, it is not clear
whether that student would need to spend another semester of mathematics in grade 8, repeat
entire lessons on PR or only focus on MVPs.

On a larger scope, TIMSS results involve a similar lack of formative use of assessment.
According to TIMSS 2007 results, the average mathematics score of U.S. eighth graders was
508, compared with all 48 participating countries on a scale average of 500. However, the 508
score provides little information about what can be done to improve students’ mathematical
knowledge. The questions remain: How much additional mathematics learning is needed and
which specific mathematics subject areas do U.S. eighth graders need help in to become at
least as proficient as students in top-performing countries such as Taiwan (598) or Korea
(597)?

An earlier set of TIMSS mathematics assessment items measured five content domains
(algebra, data, geometry, measurement, and number) and four cognitive domains (knowing
facts and procedures, using concepts, solving routine problems, and reasoning). Even when
provided with these particular measurements, one is limited in how much can be assessed
formatively. For example, one may question whether students in a country with a 508 score
need to improve their knowing facts and procedures in algebra or to spend more time solving
routine problems in geometry, or whether students in a country with a 508 score are
necessarily less proficient in using number concepts than their peers in a country with a 598
score.

MOVING TOWARD FORMATIVE ASSESSMENTS

Newly emerging psychometric frameworks, specifically CDMs, offer an alternative approach
to exploring answers to earlier questions (de la Torre, 2009; de la Torre & Lee, 2010). By
design, CDMs afford an analysis of students’ mastery or non-mastery of finer-grained sets of skills, cognitive processes, or problem-solving steps, generically referred to as attributes. Because each domain of assessment considers multiple attributes and every test item is associated with a certain combination of these attributes, one can develop a formative assessment. That is, CDMs are capable of identifying students’ strengths and weaknesses simply by examining their attribute profiles. Indeed, in contrast to an IRT framework that models the probability of an item response via a function of one examinee-item parameter, such as a single, continuous latent trait (e.g., overall mathematics proficiency), a CDM framework considers an examinee’s performance at a finer-grained level by incorporating multiple, discrete latent traits (e.g., attributes).

One CDM framework that is mostly studied in educational settings is the DINA model (deterministic inputs, noisy “and” gate). In this model, the assessment of student performance is characterized by two distinct components. The deterministic component of the DINA model classifies students into two groups based on their responses on each item: a group of students with all required attributes and another group of students lacking in at least one required attribute. The “and” gate part of this component assumes that students with all required attributes will answer the item correctly, while those lacking in at least one required attribute will answer the item incorrectly. In other words, the DINA model has a conjunctive assumption that requires mastery of all required attributes for each item.

The noise component of the DINA model allows for a stochastic process of the latent response pattern. It accounts for the probability that an examinee with all required attributes may answer the item incorrectly (slip parameter) and an examinee lacking in at least one required attribute may answer the item correctly (guessing parameter). In some cases, guessing can also be interpreted as a situation in which an examinee, despite lacking at least one required attribute, answers an item correctly by means of alternative solution methods not articulated in the list of identified attributes (de la Torre & Douglas, 2004). An extension of the DINA model, namely the Multiple-Strategy DINA (de la Torre & Douglas, 2008), discusses an approach to resolve this issue.

The CDM model has clear advantages over the IRT model given its parsimonious and formative characteristics. First, the parsimonious characteristic of the CDM model facilitates a broader range of mathematics topics to be covered in a shorter test form. For example, instead of testing three different items, each on different attribute, an examiner can test one item with a certain combination of attributes and still be able to perform a finer-grained diagnosis of the three different attributes. Second, the formative characteristic of the CDM model provides an inference about the attribute pattern of an individual student. Given the student’s test response pattern, the CDM model diagnoses the student’s mastery or non-mastery of multiple finer-grained attributes. It also provides a report about which specific mathematics skills an examinee needs to work on.
Due to the recency of the CDM framework, few analyses of large-scale assessments have been conducted with applications to the field of mathematics education. Even so, such attempts have produced drawbacks when they were naïvely executed. For instance, Tjoe, Park, and Choi (2010) could not validate to a deeper extent their underlying assumption that the TIMSS items were in fact constructed using attributes adopted from NCTM Principles and Standards. Because of this validation concern, it is critical that one develops novel items based on accurately validated attributes, rather than assigning untenably posited attributes to existing problems.

Perhaps an example of a more defensible application of CDM in the field of mathematics education was one initiated by de la Torre and colleagues (2010) on PR at the middle school level (grade 8, ages 13-14). Their study began with the search to identify finer-grained PR attributes: What is the set of cognitive processes necessary for middle school students to acquire to be deemed proficient in PR? Because a comprehensive review as done by Tourniaire and Pulos (1985) has not been completed for over two decades, de la Torre and colleagues (under review) compiled and summarized important perspectives and findings on PR research since 1985. In conjunction with this summary of conceptual and theoretical issues on PR, a group of research mathematicians, mathematics educators, and middle school mathematics teachers identified, substantiated, and verified a list of PR attributes (de la Torre & Tjoe, in preparation).

In addition to the identification process of attributes, de la Torre and Tjoe (in preparation) conducted a validation process in two parts. First, with the help of middle school mathematics teachers, the researchers created prototype PR items based on the identified attributes and verified each item-attribute specification. They did so as they considered the most common ways (among many different ways) in which middle school students would have solved these prototype PR items themselves (de la Torre & Tjoe, in preparation). Second, they interviewed middle school and college students in order to find evidence of whether students generally solved the prototype PR items in the same way they were expected to. In other words, the validation process reconfirmed that the attributes, which the researchers and middle school mathematics teachers assigned to the prototype PR items, were indeed being used by the students as they solved those items (de la Torre et al., 2012).

The six PR attributes that served as the foundation for the development of new PR items were finalized as follows: 1) prerequisite skills required in PR, 2a) comparing fractions or 2b) ordering fractions, 3a) constructing ratios or 3b) constructing proportions, 4) identifying a multiplicative relationship between sets of quantities, 5) determining whether two relations form a proportion, and 6) applying algorithms to solve a PR problem (de la Torre & Tjoe, in preparation). These attributes were grain-sized fine enough to be deemed practical for teacher diagnostic reporting and student classroom learning purposes (e.g., Attribute 4), yet coarse enough not to obscure the proposed domain of PR by compacting overly-detailed skills (e.g., Attribute 1).
At present, there is a mismatch between how these existing PR problems have been modeled by the traditional psychometric framework and how they are incapable of capturing the finer-grained cognitive processes that the newly emerging psychometric framework aimed to diagnose. As demonstrated by the findings presented in the earlier section, 69% of the existing PR problems on average were considered MVPs, while the rest were ratio-proportion construction problems and ratios comparison problems. In relation to the list of six attributes, MVPs (e.g., Item 1 in Table 1) assess Attributes 1, 3b, 5, and 6; ratio-proportion construction problems (e.g., Item 2 in Table 1) assess Attribute 3a or 3b; and ratios comparison problems (e.g., Item 3 in Table 1) assess Attributes 1, 2a or 2b, and 3a or 3b. As such, not all possible combinations of the six attributes would be assessed, thus defeating the very purpose of CDMs in evaluating students’ specific strengths and weaknesses. For example, none of the three types of PR problems measure Attributes 1 and 4.

Moreover, no inference can be made about individual attributes. Given only these three types of PR problems, any inference by CDMs about students’ mastery of a specific attribute may be confounded by the presence of other attributes. For example, one cannot use MVPs (let alone the other two types) to make a definite inference about students’ mastery of Attribute 5 without a concurrent inference about their mastery of Attributes 1, 3b, and 6. Even if a student answers MVPs correctly, it is not conclusive whether or not the student has sufficiently mastered Attribute 5 because Attribute 5 appears consistently and simultaneously in the presence of other attributes for this type of problem. On the other hand, to further consider Attribute 5, one may argue that such skill can be assessed by existing PR and non-PR problems. Yet, because there is no explicit effort to link the two different contexts of proportionality and non-proportionality in one single problem, it is not possible to find an immediate route that distinguishes between students who can differentiate a proportional situation from a non-proportional one and those who cannot.

**DESIGNING COGNITIVELY-BASED PROPORTIONAL REASONING ITEMS**

It is evident that the current state of PR problems does not effectively support the depth of the cognitive probing that CDMs were meant to model. A set of novel PR items grounded in the CDM framework is needed. To clearly distinguish these new PR items from already-existing PR problems, meaningful combinations of attributes are indispensable in developing new PR items. Without such criteria, only a mundane analysis can be conducted. This section presents two exemplars of cognitively-based PR items using the attributes validated in the earlier study by de la Torre et al. (2012).

In designing new PR items, the present researchers considered several issues. The structural or contextual components of PR problems (e.g., location of missing value, coordination of measure space, and presence of integer ratio), also need to be considered as key factors that influence the difficulty level of the PR items. In addition to this issue, de la Torre et al. (under review) discussed that certain structural or contextual components of PR problems consistently triggered students’ common misconceptions of PR (e.g., blind application of cross-multiplication procedure and inappropriate use of additive reasoning in lieu of PR). In this regard, students’ erroneous solutions can function as effective distracters in multiple-choice problems. Table 4 shows two exemplars of cognitively-based PR items.
Table 4: Two exemplars of cognitively-based PR items

Item 1: Attributes 3b and 5
Determine the appropriate proportion that can be used to solve a situation below:

Situation I: 4 pounds of tomatoes can make 7 cups of Brilliantly Yummy tomato juice. If Tommy has 12 pounds of tomatoes, how many cups of Brilliantly Yummy tomato juice can he make?

Situation II: In a classroom, there are 5 rows of chairs and there are 9 chairs in each row. If the chairs are arranged into 15 rows of chairs, how many chairs are there in a row?

\[
\begin{align*}
A. \quad & \frac{4}{7} = \frac{12}{x} \\
B. \quad & \frac{4}{7} = \frac{x}{12} \\
C. \quad & \frac{5}{9} = \frac{15}{x} \\
D. \quad & \frac{5}{9} = \frac{x}{15}
\end{align*}
\]

Item 2: Attributes 2b and 3a
Emma and Irene have 5 apples each. Penny has one less apple than Emma has. Penny and Irene have 7 cherries each. Emma has one less cherry than Penny has. Which of the following puts the order of the girls with the smallest to the greatest ratios of apples to cherries?

A. Emma, Irene, Penny
B. Irene, Emma, Penny
C. Emma, Penny, Irene
D. Penny, Irene, Emma

Item 1 in Table 4 can be thought of as a modification of MVPs. Instead of requiring students to perform such computational work as the cross-multiplication procedure, the researchers aimed to measure the students’ conceptual PR skill of differentiating situations where (direct) proportional relationships are appropriate from situations where they are not (i.e., Attribute 5). Students who decide to apply blind cross-multiplication procedures would be compelled to reflect upon the difference between options A and C. Moreover, because of the solutions involved in its multiple-choice format, Item 1 in Table 4 requires students to apply their proportion construction knowledge (i.e., Attribute 3b). In contrast to MVPs (e.g., Item 1 in Table 1), which are often associated with Attributes 1, 3b, 5, and 6, Item 1 in Table 4 can be considered unique in that it assesses a novel combination of attributes (i.e., Attributes 3b and 5). Indeed, because students are not required to perform any arithmetic computations (i.e., Attribute 1) or to apply any algorithms (i.e., Attribute 6), students’ lack of proficiency in
Attribute 1 or 6 will not confound the inference made on their mastery of attributes, which are intended to be measured for this particular item—namely, Attributes 3b and 5.

Item 2 in Table 4, on the other hand, can be thought of as a modification of ratios comparison problems. Instead of requiring students to compare two ratios only, the researchers aimed to measure students’ qualitative (rather than quantitative) PR skill of ordering three ratios by interpreting the effect of an increase or a decrease in one measure space of a ratio, keeping the other measure space constant. To this end, students need to be able to construct a ratio (albeit an abstract one) of apples to cherries for each girl (i.e., Attribute 3a), and to order the three ratios from the smallest to the greatest (i.e., Attribute 2b). With this item, one can efficiently and meaningfully combine the assessment of some of the specific attributes of interest from the ratios comparison problems (e.g., Attribute 2b of Item 3 in Table 1) and the ratio-proportion construction problems (e.g., Attribute 3a of Item 2 in Table 1). Indeed, given the level of abstraction involved in this item, Attribute 1, as commonly used by middle school students in comparing regular fractions (e.g., by finding decimal equivalents of the fractions; de la Torre & Tjoe, in preparation), has been effectively teased out from the set of attributes usually assessed in the ratios comparison problems (i.e., Attributes 1, 2a or 2b, and 3a or 3b). In short, an item with a new combination of attributes—namely, Attributes 2b and 3a—has been created.

FURTHER DIRECTIONS

A survey of the various PR problems that can be found in current mathematics textbooks, statewide examinations, and international assessments showed that on average, 69% of PR problems are typically MVPs. To optimize the usefulness of the formative aspects of CDMs, the process of designing new PR items requires conscientious deliberation over meaningful and relevant combinations of defensible attributes. The same is true for larger-scale assessments. With the increasingly growing awareness of CDMs at the international level, the field of mathematics education inevitably stands poised to become one of the major beneficiaries of these newly emerging psychometric frameworks. As a consequence, more systematic studies are needed to organize comprehensive cognitive diagnosis assessments comparable to PISA or TIMSS.

References


Tjoe & de la Torre


Lesh, R., Post, T., & Behr, M. (1988). Proportional reasoning. In J. Hiebert & M. Behr (Eds.), *Number concepts and operations in the middle grades* (pp. 93-118). Reston, VA: NCTM.


