Primitive Model of Partitive Division: A Replication of the Fischbein et al. Study

Published Version:
Primitive Model of Partitive Division: A Replication of the Fischbein et al. Study / Chiara Giberti; Andrea Maffia. - In: IMPLEMENTATION AND REPLICATION STUDIES IN MATHEMATICS EDUCATION. - ISSN 2667-0127. - ELETTRONICO. - 2:2(2022), pp. 149-173. [10.1163/26670127-bja10007]

Availability:
This version is available at: https://hdl.handle.net/11585/901508 since: 2022-11-11

Published:
DOI: http://doi.org/10.1163/26670127-bja10007

Primitive model of partitive division: a replication of the Fischbein and colleagues' study

Chiara Giberti
University of Bergamo, Italy
chiara.giberti@unibg.it
ORCID 0000-0001-7446-6709

Andrea Maffia
University of Bologna, Italy
andrea.maffia@unibo.it
ORCID 0000-0003-0080-1089

Primitive model of partitive division: a replication of the Fischbein and colleagues' study

Abstract
We present a replication study of one of the most renowned seminal works in the field of mathematics education, i.e., the Fischbein and colleagues' (1985) study on implicit models of multiplication and division in solving word problems. The original instrument was administered to a sample of 902 grade 7 students, proposing some methodological variations. The discussion focuses on four items involving partitive division of integer numbers, and results are compared to the original study: the main results of the original study are confirmed, but we contribute further with considerations on the statistical significance of differences observed in item difficulty. Finally, through our new methodological approach, we broaden the results of the original study by analysing students' behaviour on the basis of their general arithmetical problem-solving ability. Students with medium-level ability are more influenced by the primitive model than those with lower ability.

Keywords
Primitive models, division, partition, conceptual replication

1 Introduction

The teaching and learning of basic arithmetic operations is probably one of the most commonly-researched topics in mathematics education. Some of the results obtained in this sub-field of research appear to be considered as solid findings since they are cited by hundreds of *follow-up* studies. The term 'follow-up study' is here used according to its definition by Star (2021), as a study on a certain topic which draws from several prior studies related to the same topic, which also includes an innovative methodology. This kind of study seems to be very frequent in mathematics education. As observed by Star:

> This pattern - where we see many very similar follow-up studies around a given research topic, each incorporating new and often unique methodologies, assessments, and modes of analyses - appears to be the predominant norm in mathematics education research. (Star, 2021, p. 65)
>
> [...] for a given research topic, we find that there are a very small number of seminal (theoretical and empirical) papers that get cited by most studies in the subfield, while the vast majority of articles (which are not seminal but are idea-initiated follow-up studies) are almost never cited. (Star, 2021, p. 70)

One author whose work can surely be considered seminal is Efraim Fischbein (Tall, 1999). His work on the implicit models of multiplication and division published in the paper "The Role of Implicit Models in Solving Verbal Problems in Multiplication and Division" (by Efraim Fischbein, Maria Deri, Maria Sainati Nello, and Maria Sciolis Marino) in 1985 is among his most influential publications (Tirosh & Dreyfus, 1998). According to the authors of that study, "the initial didactical models seem to become so deeply rooted in the learner's mind that they continue to exert an unconscious control over mental behaviour even after the learner has acquired formal mathematical notions that are solid and correct" (Fischbein et al., 1985, p. 16). They refer to these profoundly embodied models as primitive models, asserting that repeated addition is the primitive model for multiplication while there are two primitive models for division: the partitive and the quotative. In their study, Fischbein and colleagues use these primitive models to predict the effect of changing numbers in simple word problems. For instance, in the case of partitive division, they find that for many students the dividend should be greater than the divisor in partitive division; the divisor must be a whole number; and the quotient must be smaller than the dividend.

Their results are cited in a large number of studies on this topic (almost 700 citations according to Google Scholar, which is much more than the median of major mathematics education journals according to Star, 2021), both in the past (e.g., Kouba, 1989; Graeber & Tirosh, 1990) and in very recent times (e.g. Erickson & Lockwood, 2021; Qu et al., 2021). The work by Fischbein and colleagues (1985) is cited in many relevant recent collections of research about whole number arithmetic, including the ICMI-study 23 (Chen et al., 2015) and the special issue on 'Whole number arithmetic and its teaching and learning' from the international journal ZDM (Sun, 2019). Furthermore, this paper is seminal since it introduced the well-known Fischbein's educational dilemma (Fischbein, 1987; Tirosh & Dreyfus, 1998; see next section for a description of the nature of the dilemma).

After the publication of Fischbein and colleagues' (1985) article, several researchers looked for confirmation (Kouba, 1989; Graeber & Tirosh, 1990) or confutation (Bell et al., 1989; Mulligan & Mitchelmore, 1997) of the model proposed. However, surprising as it might be, none of the aforementioned studies was an exact replication of the original study (we here clarify that we are considering a strict definition of replication study in this case). Following Star (2021), we assume that

> a replication study by design replicates most of the methodological features of the original study. Even in a conceptual replication, we would still expect that the replication study would mostly be the same, methodologically, as the original study - with only a few carefully selected areas of divergence. (p. 21)

In the case of Fischbein and colleagues' (1985) study, many have tried to replicate the results[1] but, as far as we know, nobody has replicated the study itself. As argued by several authors, a replication study can be valuable if it tests the generalizability or validity of the original study or if it reveals something that is not already known (Aguilar, 2020; Jankvist et al., 2021; Melhuish & Thanheiser, 2018; Star, 2018; 2021). We believe that, in the case of the study by Fischbein and colleagues, both these aims are feasible.

Replicating the study by Fischbein and colleagues is not an easy task because, as sometimes occurs with well-known older articles (Niss, 2019), it lacks a clear presentation of the methods. Furthermore, some methodological choices may be considered as weaknesses according to modern standards of research. Concerning the subjects involved in their research, Fischbein and colleagues (1985) describe their sample as comprising 628 pupils (228 in grade 5, 202 in grade 7 and 198 in grade 9) taken from 13 different schools in the same city, Pisa in Italy. There is no more information about the socio-economic background of those schools or whether the 13 different schools are equally distributed among the grades. The instruments used in their research are two questionnaires, each consisting of 21 items, namely short word-problems for which pupils had to indicate the solving operations (not the result). They are available in the literature, both in the original Italian version (Deri et al., 1983) and in English

[1] Examples of idea-initiated follow-up studies are provided by Graeber and Tirosh (1988), Greer (1989), Harel et al. (1994), Nesher (1987). Results of these studies are summarised in section 2.

(Fischbein et al., 1985). The rationale behind the choice of questions is explained, but no evaluation of the reliability of the instrument is provided. The 42 items designed for the experiment were divided into two questionnaires (form A and form B) and each student responded to just one. There is no evidence of equivalence of the two forms and so we cannot be sure that the comparison of items from different forms is legitimate. Also, there is no evidence of equivalence (in terms of arithmetical problem-solving ability) of the two groups of respondents.

All the aforementioned reasons point to the necessity of replicating the study to test its reliability. It might be argued that the fact that different results have never been observed (which is only partially true, see Bell et al., 1989) makes a replication study useless. While we are not convinced by this argument, we also believe there are other good reasons for replicating the Fischbein and colleagues' (1985) study. The discussion presented in the original study is based on a comparison of the percentage of correct answers provided to different items but no analysis of the significance of observed differences is provided. Smaller nuances may remain unnoticed or unprovable. The replication of the study with slightly different methods (a conceptual replication as described by Star, 2021) provides an opportunity for a fine-grained analysis. Such finer analysis requires more space to be presented and so, considering the limitations of an article, we have decided to focus only on grade 7 students as a starting point for our research (while Fischbein et al. considered grades 5, 7, and 9), and only on the primitive model of partitive division.

2 Partitive division

Multiplication and division can model several problematic situations. Researchers in mathematics education characterised these different situations according to different criteria; there is general agreement in distinguishing between partitive division and quotative (or measurement) division (Fischbein et al., 1985; Kouba, 1989; Greer, 1992; Correa et al., 1998; Squire & Bryant, 2002). According to Vergnaud (1983), when two whole numbers are multiplied, young children tend to develop a unary model of multiplication, meaning that the two factors assume different roles: there is an operator (usually the first factor in English) acting on the operand (usually the second factor in English). Based on this asymmetry in the roles of the factors, we can distinguish between the two different categories of division. According to Greer (1992, p. 276):

> Dividing the total by the number of groups to find the number in each group is called partitive division, which corresponds to the familiar practice of equal sharing (with social connotations of equity). Dividing the total by the number in each group to find the number of groups is called quotative division (sometimes termed measurement division, reflecting its conceptual links with the operation of measurement).

Research on difficulties manifested by students in selecting multiplication or division for modelling problematic situations was particularly active between the late 1980s and early 1990s, probably triggered by the publication by Fischbein, Deri, Nello, and Marino in 1985. These authors included 9 items about partitive division in their questionnaires (4 in form A and 5 in form B), which are shown in table 1. They conjectured that, when division is modelled as partitive, students will look for two specific conditions: the divisor must be smaller than the dividend and the divisor must be an integer. Also, they designed the word-problems in their questionnaire to check whether, and by how much, the level of difficulty rises when one condition or both are violated. They found a significant difference in performance when the divisor is greater than the dividend (items 16 and 17). Furthermore, results drop when the divisor is a decimal number (item 23) while there are no significant effects when the dividend is decimal. Curiously, the presence of a decimal dividend mitigates the effect of a divisor that is smaller than the dividend (items 20, 21, and 22).

Table 1. Information about partitive division items from Fischbein et al. (1985)

Items from original study	Form	Violation of the partitive model	Percentage of correct answers (grade 7)	Most common error
13) With 75 roses you can make 5 equal bouquets. How many roses will be in each bouquet?	A	None	93	75×5 (5%)
14) In 8 boxes there are 96 bottles of mineral water. How many bottles are in each box?	B	None	90	96×8 (8%)
15) I spent 1500 lire for 3 hg of nuts. What is the price of 1 hg?	B	None	89	1500×3 (5%)
16) 15 friends together bought 5 kg of cookies. How much did each one get?	A	Divisor larger than dividend	24	15÷5 (71%)
17) 12 friends together bought 5 kg of cookies. How much did each one get?	B	Divisor larger than dividend	30	12÷5 (60%)
20) To wrap 5 equal packages requires 3.25 m of string. How much string is needed for each package?	A	Divisor larger than dividend	71	5÷3.25 (13%) 3.25×5 (13%)
21) 5 friends together bought 0.75 kg of chocolate. How much does each one get?	A	Divisor larger than dividend	77	5÷0.75 (17%)
22) 5 bottles contain 1.25 L of beer. How much beer is in each bottle?	B	Divisor larger than dividend	74	5÷1.25 (17%)
23) I spent 900 lire for 0.75 hg of cocoa. What is the price of 1 hg?	B	Dividend is not an integer	25	900×0.75 (7%)

In this paper we focus our attention on the effect of the presence of a dividend that is smaller than the divisor. Having replicated the study by Fischbein and colleagues (1985), we will analyse in depth only comparable items referring to these effects in the case of whole numbers (items 13, 14, 16, and 17), since decimal numbers may mitigate the effects of the respective size of involved numbers. Several studies showed that, when writing a division operation, students order the numbers so that the larger is divided by the smaller. This is observed in purely numerical contexts (Greer, 1989) and when modelling word problems (e.g., Harel et al., 1994). The most common wrong answers shown by Fischbein and colleagues' (1985) study are consistent with this phenomenon. This phenomenon was observed with subjects of different ages, including pre-service and in-service teachers (Harel et al., 1994). We can anticipate that the difficulty of the analysed items depends on the specific numbers involved (for instance, easily detectable divisors - like 5 for 75 - may lead more easily to the choice of division as the operation to adopt) and the order of the numbers in the text of the word problem (which may increase the item difficulty when the order in the text is not consistent with the order in the solving operation, Novotná & Chvál, 2018; Searle et al., 1974; Vondrová & Novotná, 2017). However, we expect items 13 and 14 to be easier than items 16 and 17 because in items 13 and 14 the dividend is larger than the divisor; we identify these two items as 'in line with the model' of partitive division, while items 16 and 17 violate this intuitive model. According to the qualitative results presented by Graeber and Tirosh (1988), one possible explanation is provided by the fact that, in textbooks, most examples follow the 'dividend larger than the divisor' prototype. Nesher (1987) conjectures that if textbooks provided more counterexamples, students would resort more often to a logical analysis of the situation described, rather than thinking that division is always of the larger number by the smaller one. This interpretation is consistent with the belief that presenting students with situations that are inconsistent with the primitive model may help them. However, this is not the only possible position. As Fischbein and colleagues (1985) point out, teachers constantly face an educational dilemma while dealing with intuitive mathematical knowledge:

> On the one hand, if one continues to introduce the operations of multiplication and division through the models described above, one will create [...] strong, resistant, and, at the same time, incomplete models that soon will come to conflict with the formal concepts of multiplication and division. On the other hand, if one tries to avoid building the ideas related to arithmetical operations on a foundation that is behaviorally and intuitively meaningful, one certainly will violate the most elementary principles of psychology and didactics. This is one instance of a general dilemma facing mathematics teachers. (Fischbein et al., 1985, p. 15).

Furthermore, taking a position strongly depends on the source of the intuitiveness of the primitive models. As pointed out in the original article, there are two possible sources. On one hand, it is possible that intuitive models of multiplication/division reflect how these operations were initially taught in school; this provides a basis for Nesher's (1987) position. On the other hand, it might be possible that intuitive models reflect basic and natural features of human mental behaviour. As far as the authors of this paper are concerned, we do not yet have the data necessary to select one hypothesis over the other.

The fact that Fischbein and colleagues' (1985) study has been replicated – in form of conceptual replications – several times may lead to the conclusion that another replication is not necessary. However, while it is well known that difficulty in dividing a smaller number by a larger one is widespread, a deeper analysis about this kind of difficulties and about which students are more influenced by them has not yet been made. If it is true that the difficulty depends on the recurrence in textbooks of problems that do not violate the partitive model, then we should observe such difficulties in those students that (implicitly or explicitly) detect such recurrence. We may conjecture (but this should be supported by empirical data) that students who struggle more in solving arithmetical problems, might be less affected by this difficulty than those with higher arithmetical problem-solving abilities. As described in detail below, in this replication study we use the Rasch model to test this hypothesis. The research questions addressed in this study can be stated as follows:

- Considering partitive division between whole numbers, which of the differences between item difficulty (as observed by Fischbein and colleagues in the original study) appear significant when methodological improvements are implemented?
- Does the effect of the primitive model of partitive division depend on the general arithmetical problem-solving abilities of students?

While the first question finds an answer in the replication of the original study, the second one will require additional analysis of the same set of data, which we can consider a follow-up study. In the following section, we present in detail the methodological choices we made, specifying differences between the original study and our own.

3 Methods

We decided to replicate the study by Fischbein and colleagues (1985) starting from grade 7, proposing some variation at methodological level, and applying statistical methods to test the instruments and analyse the statistical relevance of differences in difficulty between the items. Starting from the same instrument used in the original work, we gathered new data by administering the test to a new sample of students; we analysed our results using the Rasch Model (Rasch, 1960), and we implemented a test-equating process (Kolen & Brennan, 2013) of the two forms.
This approach allowed us to overcome possible weaknesses of the original work and explore the results more in depth by:

- Providing a measurement of the test's reliability;
- Providing a measurement of students' ability and item difficulty;
- Comparing the difficulty of items (both in the same form or in different forms) and analysing whether differences between items are statistically significant;
- Analysing students' answers to a specific item, considering their ability across the whole test.

The number of cases allowed us to explore and compare the functionality of items using the Rasch Model (Rasch, 1960), which is the simplest model belonging to the Item Response Theory (IRT) and which assumes that the probability for a student to provide a correct answer to an item is governed by his/her relative ability, i.e., the intrinsic ability of the student compared to the difficulty of the item. The latent trait measured by the test (to which we refer, in our case, as arithmetical problem-solving ability) has been estimated by the Rasch model and scaled in an empirical range equal to [-4, +4]. We thus had the estimation of item difficulty (Delta) and students' ability on the same scale: negative values correspond to easier items and students with low ability levels, while positive values correspond to difficult items and students with high ability levels. Rasch's analysis made it possible to compare not only the functionality of each item, but also the functionality of the tests as a whole. Indeed, in their work, Fischbein and colleagues (1985), explained in depth the rationale behind the choice of questions, but did not report any measurement of the reliability of the instrument. The Rasch model was implemented using the jMetrik software, which also adds some measures from the Classical Test Theory (CTT, such as Discrimination index and Cronbach's alpha). The reliability of the instrument (i.e., the consistency of the measurement, the reproducibility of the results) was measured in terms of Guttman's L2 coefficient and Cronbach alpha (Meyer, 2014). The analysis computed through the Rasch model accompanied by some of the main indexes of the CTT are consistent with previous studies and with the analysis performed for national surveys by the Italian National Institute for the Evaluation of Educational System (i.e., Bolondi et al., 2018; Istituto nazionale per la valutazione del sistema di istruzione e formazione [INVALSI], 2012). Furthermore, even though IRT models overcome many typical obstacles of CTT models (Rasch, 1960), some difficulties emerge in the calculation of reliability when dealing with binary responses; thus, approximate reliability measures, such as Cronbach's alpha, are preferred when using IRT models (Milanzi et al., 2015).

We administered Fischbein and colleagues' (1985) test in a computer-based version. This has led to two important improvements from a methodological point of view, allowing us to reach schools from different geographical regions and involve a large number of students from different backgrounds, as well as completely randomising the order of questions in the tests.

The test in the original work was divided into two forms (A and B), both of which contained the same number of items and the same distribution between multiplication, division and addition/subtraction items. As in the original research study, the two forms were distributed randomly to the students, but this does not directly guarantee comparability of the two groups of students or of their answers. Hence, we implemented a test-equating procedure (Kolen & Brennan, 2013) based on the Rasch Model (Rasch, 1960), to ensure direct comparison of the items included in the two forms. We then used unvaried items (included in both forms) as an anchor, allowing us to compare the results of the two groups of students. Thanks to this methodological improvement, we also had the possibility of locating all the items from the two forms on the same difficulty scale (based on the Rasch difficulty parameter, Delta) and comparing similar items included in the two forms, taking into account the fact that different students responded to the two forms. Furthermore, the test-equating procedure shows whether these differences between item difficulties are statistically significant or not, considering their difficulty parameter, standard error, and a confidence level of 95%. Following this procedure, our first research question can be re-state as follow: for which items about partitive division do we observe statistically significant differences in results? We expect that items violating the intuitive model of partitive division (items 16 and 17) have a significantly lower score than those not violating the intuitive model (items 13 and 14).

Finally, we analyse a specific chart for each item of the test (output of the Rasch model) called a 'distractor plot'. A distractor plot is a Cartesian plot which reports the probability of choosing a specific answer (correct or incorrect) in relation to students' ability across the whole test. We draw distractor plots dividing students into five quintiles, on the basis of their arithmetical problem-solving ability measured across all items in the test. We plot a curve for each possible label (correct answer, wrong numbers, wrong operation, inversion of terms, as listed below in section 3.3): each curve shows the percentage of students from each quintile who provide a certain kind of answer. For each quintile, we plot dots whose abscissa corresponds to the average ability level of the quintile. The distractor plot analysis allows us to study students' most common answers while considering their arithmetic ability level, so providing an answer to our second research question.

3.1 Sample

In this paper we present the results of our replication of the Fischbein and colleagues' study, considering only grade 7. In further research we will also include grade 5 and grade 9, as in the original research. Beginning from grade 7 appeared the best choice because, whatever the next step is, it will be easier to compare the results with a gap of two years (between grade 7 and 5, or between grade 7 and 9) instead of a gap of four years that would be present if we started with grade 5 and then could only move to grade 9 next. Also, from an organisational point of view (in terms of availability of national data used to construct the sample), starting from grade 7 was possible. We involved 22 schools from different parts of Northern Italy; in each school, 2 or 3 classes participated in the study and we thus reached 46 classes in total. Schools were chosen on the basis of their socio-economic background (considering information gathered by the Italian National Institute of Evaluation of the Educational System and/or directly by the schools) in order to ensure the greatest possible variety of backgrounds. Our sample is composed of 902 grade 7 students (485 responded to form A and 417 to form B); the sample of the original study, considering only grade 7, included 202 students from schools of the same Italian city (Pisa) and no background information was made explicit in the study.

3.2 Instruments

We started from the test used by Fischbein and colleagues in the original research which is available both in Italian and English languages (for the Italian version, see Deri et al., 1983; for English, see Fischbein et al., 1985). The fact that the original research was implemented in Italian schools allowed us to use the instrument in its original version and no translation was required.

The original test was composed of 42 items: 12 multiplication problems, 14 division problems, and other 16 subtraction or addition problems; these latter problems were not the focus of the research but reduced the likelihood of fortuitous correct answers. The test then measured students' ability in arithmetic problem solving. With the aim of reducing the fatigue effect, the 42 items were divided by the original authors into two forms composed of 21 items each (6 multiplication problems and 7 division problems were included in each form), which were randomly distributed in each classroom.

We started from the original instrument, maintaining the multiplication and division problems unvaried and with the same partitioning into the two forms (we called these two forms A and B as in the original study); thus, items 13 and 16 were included in form A, while items 14 and 17 were included in form B. To implement the test-equating procedure we decided to form the anchor between the two forms maintaining the same unvaried addition and subtraction items (we used the items of one of the two forms of the original instrument). Furthermore, we added four new unvaried items (concerning multiplication and division) to improve the robustness of the anchor. Finally, on the basis of a pilot study, we implemented small variations to the wording in a few of the problems in order to bring them closer to students' current experience and language[2].

In the original study, to reduce the order effect, questions were ordered in half of the booklets in one way, and in reverse order for the other half. Analysing their data, Fischbein and colleagues (1985) affirmed that no order effect was found. In our study, the order of all the questions is completely randomised and this allows us to state with even higher certainty that the order of the questions did not affect students' answers.

3.3 Data analysis

[2] No modification in question formulation was due to the translation of the text because we used the questions included in a paper published in Italian by the original authors (Deri et al., 1983). A few changes were implemented to overcome possible obstacles due to words which are no longer frequently used in the Italian daily language (e.g. we substitute the Italian word 'dono' with the more common 'regalo', both the words mean 'gift'; we changed 'garofani' (carnations) with 'rose' (roses) because, during the pilot study, many students asked the about the meaning of this word). Furthermore, we performed minor changes in the writing of the units of measure, avoiding abbreviation (e.g., 'kilograms' was chosen instead of 'kg').

We collected data and classified students' answers as correct, missing or wrong. We labelled wrong answers according to the nature of the mistake, considering the most common types of wrong answers found in the original study:

- Inversion of the order of terms (e.g., 12:5 instead of 5:12)
- Wrong numbers (e.g., 5:13 instead of 5:12)
- Reverse operation (e.g., 5×12 instead of 5:12)
- Other non-equivalent operations.

The first step of data analysis consists in the analysis of percentages of each type of answer, to enable a direct comparison of our results with those of Fischbein and colleagues (1985).
The second step of data analysis was implemented with the software jMetrik. We applied the Rasch Model to the whole dataset, specifying which items were included only in form A, which ones only in form B and which items formed the anchor (included in both forms). In this way, the test-equating was performed, allowing us to locate all the students on the same ability scale, and all the items on the same difficulty scale. According to the Rasch model, both item difficulty (Delta) and students' arithmetical problem-solving ability (measured across the whole test) are placed on the same scale with values between -4 and +4, where -4 is the lowest ability/difficulty level, while +4 is the maximum ability/difficulty.
The reliability analysis after the anchoring technique highlights good internal consistency in terms of Guttman's L2 coefficient (0.84) and acceptable in terms of Cronbach alpha (0.75). Furthermore, high reliability and separation, in terms of both persons and items, indicate that the sample is large enough to consistently rank items by difficulty and respondents by ability (Meyer, 2014).
The quality of the instrument, and in particular of the anchor, is also confirmed by the distribution of items in relation to their difficulty after the test-equating procedure (Fig. 1): the majority of the items in the two forms are concentrated in the centre of the scale covering the latent trait for values between -3 and +3; the anchoring items (green dots) are differentiated in terms of difficulty and span almost the entire scale. The items about partitive division that we are analysing are at the endpoints of the interval.

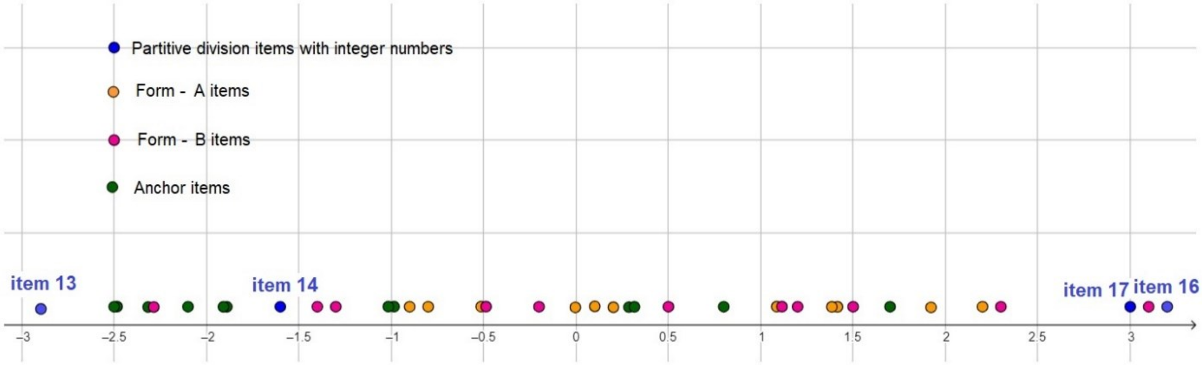


Figure 1. Distribution of the difficulty of items (based on Rasch Delta parameter) considering the two forms after the test-equating procedure. Blue dots are the partitive division items analysed in this study.

4 Results and Discussion

In this paper, we focus on the results of four items involving partitive division with whole numbers (Table 1). To improve readability of the results, we labelled these items considering the contexts and the correct answers, as follows:

- Item 13 - *Roses Item [75÷5]*;
- Item 14 - *Bottles Item [96÷8]*;
- Item 16 - *Cookies Item [5÷15]*;
- Item 17 - *Cookies Item [5÷12]*.

In each of the two forms of the test, we have one standard partitive division word-problem (*Roses Item [75÷5]* and *Bottles Item [96÷8]*) and approximately 90% of the students give the correct answer to these two problems, confirming Fischbein's results (Table 2). Also, each form included a word-problem with the divisor larger than the dividend (*Cookies Item [5÷15]* and *Cookies Item [5÷12]*), so violating the primitive model of partitive division. The percentage of correct answers drops for these two items both in the original research study (24% in *Cookies Item [5÷15]* and 30% in *Cookies Item [5÷12]*) and, even more so, in ours (16% in *Cookies Item [5÷15]* and 19% in *Cookies Item [5÷12]*). Then, the violation of the primitive model of partitive division has a strong influence on students' answers and this is evidenced also by the incorrect answers given in the two studies: in both items (*Cookies Item [5÷15]* and *Cookies Item [5÷12]*) approximately 70% of students answer with the right operation (division), but with an inversion of the order of terms, so reporting an operation which is in line with the primitive model of partitive division. Our study confirms Fischbein's results and, in our research, the drop in correct answers between the two standard items and those violating the primitive models, is even stronger than in the Fischbein and colleagues' study (Table 2).

Table 2. Comparison of our and Fischbein et al. (1985) results in partitive division items

	Item's features					Fischbein et al. (1985) results at grade 7		Our results at grade 7			
Items involving partitive division	Correct answer	Most common error	In line with the primitive model	Identification of multiples/divisors	Order of the number in the text	Correct answer	Most common error	Correct answers	Most common error	Delta	St. error
Roses Item [75÷5] 13) With 75 roses you can make 5 equal bouquets. How many roses will be in each bouquet?	75÷5	75×5	yes	Easy and helpful for the resolution	Consistent with the solving operation	93%	5%	93%	3%	-2.94	0.24
Bottles Item [96÷8] 14) In 8 boxes there are 96 bottles of mineral water. How many bottles are in each box?	96÷8	96×8	yes	Medium difficulty and helpful for the resolution	Not consistent with the solving operation	90%	8%	88%	7%	-1.58	0.17
Cookies Item [5÷15] 16) 15 friends together bought 5 kg of cookies. How much did each one get?	5÷15	15÷5	no	Easy but hindering for the resolution	Not consistent with the solving operation	24%	71%	16%	73%	3.21	0.14
Cookies Item [5÷12] 17) 12 friends together bought 5 kg of cookies. How much did each one get?	5÷12	12÷5	no	No multiples/divisors	Not consistent with the solving operation	30%	60%	19%	72%	3.03	0.15

The results in terms of percentages of correct and incorrect answers allow a direct comparison between our results and those reported in the original research. Despite this, the comparison of item difficulty and, in particular, the comparison between items included in different forms, is more accurate on the basis of the test-equating procedure (last two columns in Table 2). Indeed, using the anchoring technique, we can directly compare items' difficulty because all the items of the two forms are located on the same difficulty scale by the Rasch model: the two items violating the primitive model of partitive division (*Cookies Item [5÷15]* and *Cookies Item [5÷12]*) are significantly more difficult than the other two (*Roses Item [75÷5]* and *Bottles Item [96÷8]*) and we notice a change in the difficulty parameter from -2.9/-1.6 of the first two items to +3.2/+3.0 of the second ones (Delta index reported in Table 2) which is statistically significant. The results of the test-equating also confirm the different difficulty within these two pairs of items: *Roses Item [75÷5]* is easier than *Bottles Item [96÷8]* while *Cookies Item [5÷12]* is easier than *Cookies Item [5÷15]* (Figure 1 and Table 2). Considering the standard error of each difficulty parameter, we found that the difference, in terms of difficulty parameters, between *Roses Item [75÷5]* and *Bottles Item [96÷8]* is statistically significant, unlike the difference between *Cookies Item [5÷15]* and *Cookies Item [5÷12]*, which is not so. Then, considering the two items in line with the primitive model (*Roses Item [75÷5]* and *Bottles Item [96÷8]*), we observe that the higher difficulty of *Bottles Item [96÷8]* might be due to the order of the numbers in the text, which is not consistent with the order of the solving operation (as already observed by Novotná & Chvál, 2018; Searle et al., 1974; Vondrová & Novotná, 2017); furthermore, in both these items the two involved numbers are one multiple of the other (leading to integer results), but it is easier to identify 5 as a divisor of 75 rather than 8 as a divisor of 96, and this might help students in *Roses Item [75÷5]*. This last consideration could also influence students' answers to *Cookies Item [5÷15]* and *Cookies Item [5÷12]*: indeed, in the former, the two numbers are 15 and 5, and 5 could be easily identified as a divisor of 15, while in the latter the numbers are 12 and 5. The higher difficulty of *Cookies Item [5÷15]* might be connected to the fact that an easily detectable divisor immediately leads to the choice of a division with the divisor smaller than the dividend; this is evident in the results of Fischbein and colleagues' (1985), where the percentage of students inverting the two numbers is higher for *Cookies Item [5÷15]* (71% in *Cookies Item [5÷15]* and 60% in *Cookies Item [5÷12]*), but this is not so evident if we consider the same error in our study (approximately 70% in both items). Finally, in these two items, the order of numbers in the text is not consistent with the order in the solving operation (Novotná & Chvál, 2018; Searle et al., 1974; Vondrová & Novotná, 2017) and this probably increases their difficulty.

We can state that a strong (in the sense of statistically significant) difference in terms of item difficulty emerges between items violating the primitive model of partitive division and items in line with the model. Moreover, the order of factors in the text has an impact on students' answers, helping them to identify the correct operation if the order is coherent (*Roses Item [75÷5]*) or becoming an obstacle if it is not (*Bottles Item [96÷8]*); this obstacle could be even greater if the item violates the primitive model and the divisor is higher than the dividend (*Cookies Item [5÷15]* and *Cookies Item [5÷12]*). In all items, except *Cookies Item [5÷12]*, the two numbers are multiples, one of the other; also in this case, the ease of identifying a divisor could help in some situations (items conforming to the primitive model of partitive division, e.g., *Roses Item [75÷5]*) and become an obstacle in others (items violating the primitive model, e.g. *Cookies Item [5÷15]*). Even though 96 is a multiple of 8, it is not an easily detectable divisor and this could be a possible additional reason for the higher difficulty of this item as compared with *Roses Item [75÷5]*.

The test-equating technique based on the Rasch model allows us to analyse the distractor plot of each item: the distractor plot shows the trend of students' answers (correct, wrong, and missing answers) as a function of student ability across the whole test (in this case, arithmetical problem-solving ability). The x-axes of each plot represent the students' ability level (Delta) measured by the test and directly comparable thanks to the test-equating technique; the y-axes reports, for each quintile, the percentage of students answering correctly (green dots), incorrectly (considering different type of incorrect answers), and the percentage of missing answers (yellow dots). We observe that, *as expected, in the two tasks that are in line with the primitive model (Roses Item [75÷5] and Bottles Item [96÷8]*, Fig. 2) the correct answer (green lines) has an increasing trend and, from medium ability (Delta=0) onwards, almost all the students answer correctly. Considering lower ability levels (Delta<−1), in both items the most common errors are connected to the choice of an incorrect operation and, in particular in *Bottles Item [96÷8]*, a significant percentage of students wrote the product of the two numbers in the text (96×8, pink dots, approximately 20% of students from the first quintile).

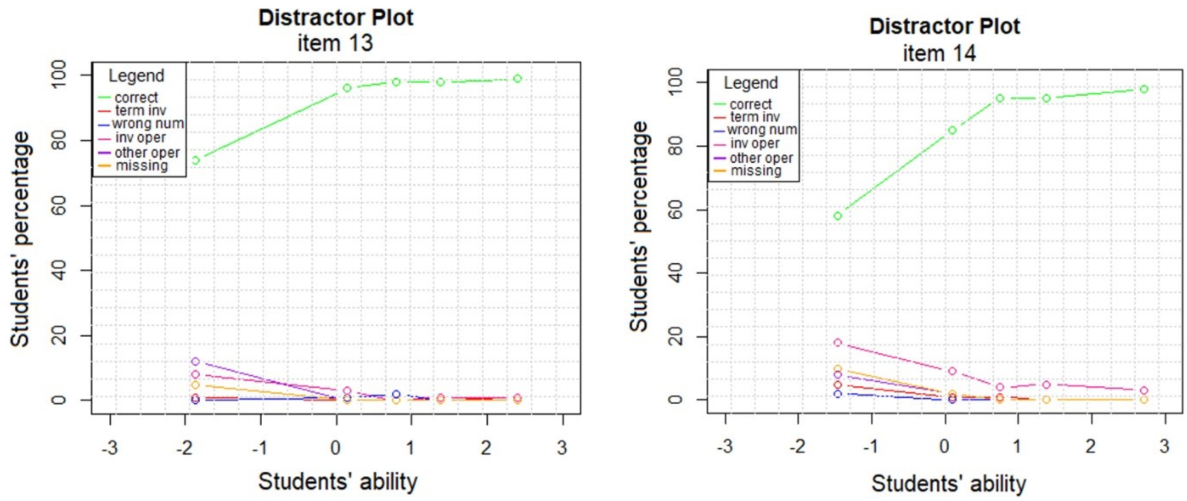
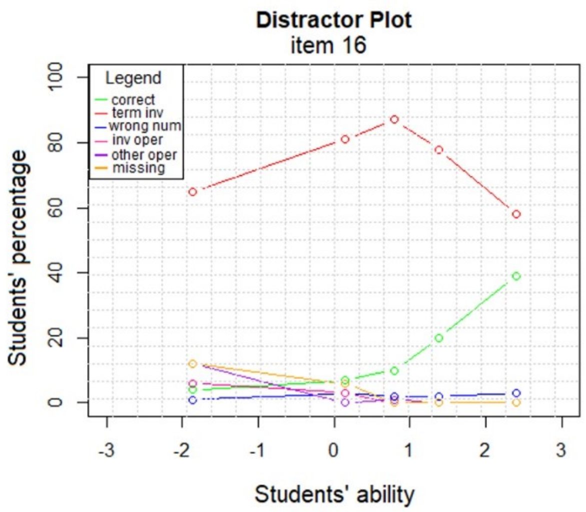


Figure 2. Distractor plots for *Roses Item [75÷5]* (left plot) and *Bottles Item [96÷8]* (right plot). On the x-axis we report students' arithmetical problem-solving ability measured over the entire test; on the y-axis we report, for each quintile, the percentage of students for each answer (correct answers, different types of wrong answers, and missing answers).

The most interesting finding emerges in the analysis of the distractor plots of *Cookies Item [5÷15]* and *Cookies Item [5÷12]* (Fig. 3). The trend of the correct answer is, as expected, increasing, but the growth is weaker for low (Delta<−1) and medium ability levels, before becoming higher from medium levels (Delta>1) upward. A large majority of students of all levels answers these items by selecting the correct operation (division), but with the terms inverted (red lines). The trend of this kind of wrong answer is peculiar: it increases from lower (Delta<−1) to medium ability levels and then it decreases. We observe this trend in the distractor plot on the left in relation to *Cookies Item [5÷15]:* the percentage of students answering with the inversion of the terms (15÷5) is approximately 70% in the first quartile (the one representing lowest ability levels), then this percentage increases in the second and third quartile (respectively 80% and 90%), while decreasing for the fourth quintile (80%) and even more so in the highest ability quintile (60%). In the distractor plot on the right, related to *Cookies Item [5÷12],* the two numbers are not multiples one of the other, but the trend observed for the analogous incorrect answer (inversion of the terms,12÷5) is similar: the percentage of students answering with inversion of the terms is approximately 60% for the lowest and the highest quintile, while the percentage is higher in the second, third, and fourth quintile (respectively 85%, 75%, and 75%), thus for medium ability levels. This is a very particular trend for an incorrect answer because, typically, the expected trend of an incorrect answer is a complete decrease: thus, students with lower ability levels measured over the entire test have a higher probability of choosing an incorrect answer and, as their ability grows, the incorrect answer becomes less attractive. This kind of trend does not represent a violation of the Rasch Model and was already observed in the literature: Ferretti and colleagues (2018) named it the ‘humped performance trend’. They observed this trend for items in which an incorrect answer was strictly linked to obstacles due to didactic practices, and in particular to misconceptions and didactic contract (Brousseau, 2002) effects. Indeed, students with medium ability are typically students who are strongly conditioned by classroom routines, didactic practices, and instruction given by the teacher (Ferretti et al., 2018). For this reason, they struggle more to overcome obstacles due to implicit rules or implicit wrong models which emerge in the relationship between the teacher and students. For instance, Ferretti and colleagues (2018) observe this trend in relation to items in which a selective reading of the text, which is frequently encouraged by widespread didactical practices (i.e., the request to circle all numerical data and underline the question; Vondrová & Novotná, 2017), leads students to consider only numerical data; thus, they observe that this kind of error is more frequent for students with medium ability levels. Also in our results, we observe that students with medium ability levels (thus, students who should be more aware of didactic practices), are particularly influenced by the primitive model of division: in both the two items violating the primitive model of partitive division (*Cookies Item [5÷15]* and *Cookies Item [5÷12]*) they identify the correct operation (division), but they are far less able to overcome the implicit rule following which a division has the dividend larger than the divisor (thus answering respectively 15÷5 and 12÷5). As mentioned by some authors (Graeber & Tirosh, 1988; Nesher, 1987), this implicit rule (and the difficulty in solving problems violating the division primitive model) may be rooted in the didactic practices of the teachers and the examples that students encountered in the past (including examples provided by textbooks). The distractor plots highlight that, in both items, this difficulty is also evident for the top performing students: considering the highest percentile (Delta>2), we observe that the percentage of students who choose the correct answer (green line, approximately 40% in both items) is significantly lower than the percentage of students answering with the division but inverting the two terms (red line, approximately 60% in both items).



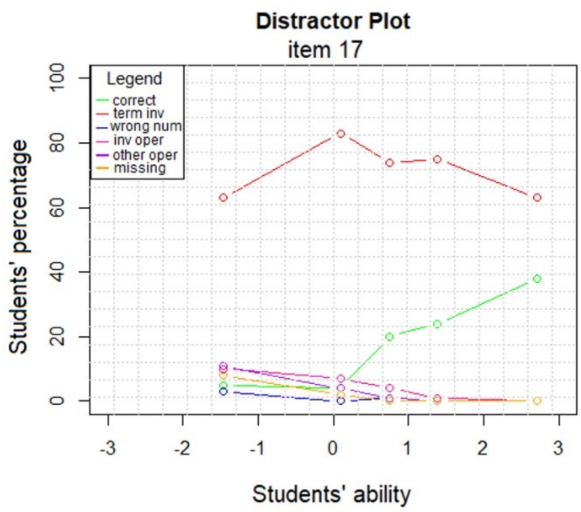


Figure 3. Distractor plots for *Cookies Item [5÷15]* and *Cookies Item [5÷12]*. On the x-axis we report students' arithmetical problem-solving ability measured over the entire test; on the y-axis we report, for each quintile, the percentage of students for each answer (correct answer, different types of wrong answers, and missing answers).

5 Conclusion

Replication studies are relevant when they test the generalizability or validity of the original study or if they let us learn something new (Aguilar, 2020; Jankvist et al., 2021; Melhuish & Thanheiser, 2018; Star, 2018; 2021). We believe that, by answering our research questions, we contributed in both directions.

Firstly, by using the same items designed by Fischbein and colleagues (1985), we added information about the instruments. The original study provided detailed information about the design of the items, guaranteeing the validity of the instruments. By repeating the study, we added information about their reliability. Adopting a test-equating technique and through an actual randomization of items presentation, we also improved the method. These improvements were possible also because we adopted a computer-based method of administration that was not possible when the original study was conducted. Therefore, we can observe that repetition of classic studies may be important in providing the research field with updated and robust instruments that can still be used in more modern research.

Thanks to the anchoring technique, the results obtained from our study add additional information to the original research by attesting the significance of the difference in difficulty between the items on which we focused. The two items violating the intuitive model of partitive division were significantly more difficult than the two items that were consistent with the model. This is consistent with the strong difference in percentage of correct answers observed in the original study. However, we also noticed that the difficulty of the *Roses Item [75÷5]* and *Bottles Item [96÷8]* (both with the dividend greater than the divisor) is significantly different. A possible explanation is offered by the fact that, in one case, the second term of the division is more easily identified as a divisor of the first term. Another difference is given by the order of appearance of the numbers in the text of the word-problem: the case in which the order in the text is consistent with the order in the division, results as easier. This was already observed in the literature (Novotná & Chvál, 2018; Searle et al., 1974; Vondrová & Novotná, 2017). This difference in the difficulty of *Roses Item [75÷5]* and *Bottles Item [96÷8]* (that was not apparent in the original study) appears extremely relevant since the possible interpretations may also explain part of the difficulty of the *Cookies Item [5÷15]* and *Cookies Item [5÷12]*; thus, the effect of the intuitive model on the difficulty of the item may be increased by the choice of the numbers and/or by their order in the text. Specific research could be devoted to confirming this conjecture. For instance, it would be interesting to see how students respond to an item in which the primitive model is violated, but the order of the numbers in the text corresponds to the order of the numbers in the division (there is no item with these characteristics in the original study).

Comparing our study to the original research, we can also note that we obtained a lower percentage of correct answers for items violating the primitive model of partitive division, and a higher percentage of inversion of the two terms. These differences between our study and that of Fischbein and colleagues' might be due to the period of time which lapsed between the two studies (in Italy, there was a considerable curriculum reform in the interim), but also to the fact that our sample is larger and more stratified in terms of socio-economic background. It is impossible to address the cause based on the currently available data. To better understand the effects of curriculum changes or of characteristics of the sample of this kind of result, further replications in different contexts are needed.

By modifying the methods of the original study (using the Rasch model and a test-equating technique) we were also able to produce a finer-grain analysis of students' answers, comparing the difficulty they encountered in the selected items with their ability demonstrated across the whole test. We found that the probability of inverting the order of terms in a division with the divisor greater than the dividend, is higher for students with a medium-level ability than for those with low ability (humped performance trend, Ferretti et al., 2018). Since this trend is often observable for difficulties due to didactic practices, our results substantiate quantitatively what was previously suggested qualitatively (Graeber & Tirosh, 1988; Nesher, 1987). While we are still far from resolving the doubt about the source of intuitive models (i.e., whether they are caused only by educational practices or are the effect of an innate cognitive disposition), our data seem to reinforce the hypothesis of didactic causes.

The new information provided by this repetition of the Fischbein and colleagues' (1985) study on primitive model prompts us to propose the replication also for the other two grades considered in the original study (5 and 9). More generally, we believe that the relevance of the results presented in this contribution may serve as a trigger to repeat other important seminal studies in the field of mathematics education.

Acknowledgement

The authors wish to thank the Principals and teachers who collaborated in the data collection during the difficult times of pandemic. Without their collaboration, this study would have not been possible. We would like to sincerely thank Dr. Clelia Cascella for her advice and collaboration in defining the statistical methods for the present study.

References

Aguilar, M. S. (2020). Replication studies in mathematics education: What kind of questions would be productive to explore? *International Journal of Science and Mathematics Education*, *18*(1, Suppl.), S37–S50. https://doi.org/10.1007/s10763-020-10069-7

Bell, A., Greer, B., Grimison, L., & Mangan, C. (1989). Children's performance on multiplicative word problems: Elements of a descriptive theory. *Journal for Research in Mathematics Education*, *20*(5), 434–449. https://doi.org/10.5951/jresematheduc.20.5.0434

Bolondi, G., Cascella, C., & Giberti, C. (2018). How much item formulations affect the probability of a correct answer? An experimental study. In Curtis, F. (Ed.), *Proceedings of the British Society for Research into Learning Mathematics*, *38*(2). https://www.bsrlm.org.uk/publications/proceedings-of-day-conference/ip38-2/

Brousseau, G. (2002). *Theory of didactical situations in mathematics. Didactique des mathématiques 1970–1990* (N. Balacheff, M. Cooper, R. Sutherland, & V. Warfield, Eds. & Trans.). Springer. https://doi.org/10.1007/0-306-47211-2

Chen, L., Van Dooren, W., Jing, M., & Verschaffel, L. (2015). Effect of learning context on students' understanding of the multiplication and division rule for rational numbers. In X. Sun, B. Kaur, & J. Novotna (Eds.), *Conference proceedings of ICMI study 23: Primary mathematics study on whole numbers* (pp. 526–533). ICMI.

Correa, J., Nunes, T., & Bryant, P. (1998). Young children's understanding of division: The relationship between division terms in a noncomputational task. *Journal of Educational Psychology, 90*(2), 321–329. https://doi.org/10.1037/0022-0663.90.2.321

Deri, M., Nello, M., & Marino, M. (1983). Il ruolo dei modelli primitivi per la moltiplicazione e la divisione [The role of primitive models for multiplication and division]. *L'insegnamento della matematica e delle scienze integrate, 6*(6), 6–27.

Erickson, S.A., & Lockwood, E. (2021). Investigating combinatorial provers' reasoning about multiplication. *International Journal of Research in Undergraduate Mathematics Education*, *7*(1), 77–106. https://doi.org/10.1007/s40753-020-00123-8

Ferretti, F., Giberti, C. & Lemmo, A. (2018). The Didactic Contract to interpret some statistical evidence in mathematics standardized assessment tests. *EURASIA Journal of Mathematics, Science and Technology Education*, *14*(7), 2895–2906. https://doi.org/10.29333/ejmste/90988

Fischbein, H. (1987). *Intuition in science and mathematics: An educational approach*. Springer.

Fischbein, E., Deri, M., Nello, M. S., & Marino, M. S. (1985). The role of implicit models in solving verbal problems in multiplication and division. *Journal for Research in Mathematics Education*, *16*(1), 3–17. https://doi.org/10.5951/jresematheduc.16.1.0003

Graeber, A., & Tirosh, D. (1988). Multiplication and division involving decimals: Preservice elementary teachers' performance and beliefs. *The Journal of Mathematical Behavior, 7*, 263–280.

Graeber, A. O, & Tirosh, D. (1990). Insights fourth and fifth graders bring to multiplication and division with decimals. *Educational Studies in Mathematics*, *21*(6), 565–588. https://doi.org/10.1007/BF00315945

Greer, B (1989). Conceptual obstacles to the development of the concepts of multiplication and division. In H. Mandl, E. de Corte, S. N. Bennet, & H. Friedrich (Eds.), *Learning and instruction: European research in an international context.* (Vol. 22, pp. 461–476). Pergamon.

Greer, B. (1992) Multiplication and division as models of situations. In Grouws, D. A. (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 276–295). Macmillan.

Harel, G., Behr, M., Post, T., & Lesh, R. (1994). The impact of the number type on the solution of multiplication and division problems: Further considerations. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 365–384). State University of New York.

INVALSI. (2012). *Rilevazione nazionale sugli apprendimenti 2011–12* [National survey on learning 2011-12]. https://www.invalsi.it/snv2012/documenti/Rapporti/Rapporto_tecnico_SNV2012.pdf

Jankvist, U. T., Aguilar, M. S, Misfeldt, M., & Koichu, B. (2021). Launching Implementation and Replication Studies in Mathematics Education (IRME). *Implementation and Replication Studies in Mathematics Education*, *1*(1), 1–19. https://doi.org/10.1163/26670127-01010001

Kolen, M. J., & Brennan, R. L. (2013). *Test equating: Methods and practices*. Springer.

Kouba, V. L. (1989). Children's solution strategies for equivalent set multiplication and division word problems. *Journal for Research in Mathematics Education*, *20*(2), 147–158. https://doi.org/10.5951/jresematheduc.20.2.0147

Meyer, J. P. (2014). *Applied measurement with jMetrik.* Routledge.

Melhuish, K., & Thanheiser, E. (2018). Reframing replication studies as studies of generalizability: A response to critiques of the nature and necessity of replication. *Journal for Research in Mathematics Education*, *49*(1), 104–110. https://doi.org/10.5951/jresematheduc.49.1.0104

Milanzi, E., Molenberghs, G., Alonso, A., Verbeke, G., & De Boeck, P. (2015). Reliability measures in item response theory: Manifest versus latent correlation functions. *The British Journal of Mathematical and Statistical Psychology*, *68*(1), 43–64. https://doi.org/10.1111/bmsp.12033

Mulligan, J. T., & Mitchelmore, M. C. (1997). Young children's intuitive models of multiplication and division. *Journal for Research in Mathematics Education*, *28*(3), 309–330. https://doi.org/10.5951/jresematheduc.28.3.0309

Nesher, P. (1987). Towards and instructional theory: The role of student's misconceptions. *For the Learning of Mathematics*, *7*(3), 33–40.

Niss, M. (2019). The very multi-faceted nature of mathematics education research. *For the Learning of Mathematics*, *39*(2), 2–7.

Novotná, J., & Chvál, M. (2018). Impact of order of data in word problems on division of a whole into unequal parts. *Journal on Efficiency and Responsibility in Education and Science*, *11*(4), 85–92. https://doi.org/10.7160/eriesj.2018.110403

Qu, C., Szkudlarek, E., & Brannon, E. M. (2021). Approximate multiplication in young children prior to multiplication instruction. *Journal of Experimental Child Psychology, 207*, Article 105116. https://doi.org/10.1016/j.jecp.2021.105116

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danmarks Paedagogiske Institut.

Searle, B. W., Lorton, P., & Suppes, P. (1974). Structural variables affecting CAI performance on arithmetic word problems of disadvantaged and deaf students. *Educational Studies in Mathematics*, *5*(1), 371–384. https://doi.org/10.1007/BF00684708

Star, J. R. (2018). When and why replication studies should be published: Guidelines for mathematics education research journals. *Journal for Research in Mathematics Education*, *49*(1), 98–103. https://doi.org/10.5951/jresematheduc.49.1.0098

Star, J. R. (2021). In pursuit of a replication culture in mathematics education. *Implementation and Replication Studies in Mathematics Education,* *1*(1), 53–76. https://doi.org/10.1163/26670127-01010003

Squire, S., & Bryant, P. (2002). From sharing to dividing: Young children's understanding of division. *Developmental Science*, *5*(4), 452–466. https://doi.org/10.1111/1467-7687.00240

Sun, X. H. (2019). Bridging whole numbers and fractions: problem variations in Chinese mathematics textbook examples. *ZDM – Mathematics Education, 51*(1), 109–123. https://doi.org/10.1007/s11858-018-01013-9

Tall, D. (1999). *Efraim Fischbein, 1920-1998, Founder President of PME: A Tribute*. https://homepages.warwick.ac.uk/staff/David.Tall/pdfs/dot1999b-fischbein-tribute.pdf

Tirosh, D, & Dreyfus, T. (1998). Efraim Fischbein, 1920-1998. In Niss, M (Ed.) *ICMI Bullettin, n. 45*. https://www.mathunion.org/fileadmin/IMU/Organization/ICMI/bulletin/45/Fischbein.html

Vergnaud, G. (1983) Multiplicative structures. In L. Landau (Ed.), *Acquisition of mathematics concepts and processes* (pp. 127–174). Academic Press.

Vondrová, N., & Novotná, J. (2017). The influence of context and order of numerical data on the difficulty of word problems for grade 6 pupils. In J. Novotná & H. Moraová (Eds.), *International Symposium Elementary Maths Teaching: Equity and diversity in elementary mathematics education* (Vol. 26, pp. 440–448). Charles University.