BRILL

IMPLEMENTATION AND REPLICATION STUDIES IN MATHEMATICS EDUCATION 3 (2023) 1–43

IRME

brill.com/irme

# Danish Students' Understanding of Fractions: A Replication Study

*Julie Vangsøe Færch* | ORCID: 0009-0003-7989-7438
University College Copenhagen, Humletorvet 3,
1799 Copenhagen V, Denmark
Department of Educational Theory and Curriculum Studies.
Danish School of Education, Aarhus University, Nobelparken,
Building 1483, Jens Chr. Skous Vej 4, 8000 Aarhus C, Denmark
*jvaf@kp.dk; jvaf@edu.au.dk*

*Jeremy Hodgen* | ORCID: 0000-0002-9196-4088
University College London, Gower Street, London, WC1E 6BT, UK
*Corresponding author, e-mail: jeremy.hodgen@ucl.ac.uk*

Received 15 July 2023 | Accepted 12 September 2023 |
Published online 9 November 2023

**Abstract**

In this paper, we report on a conceptual replication of a fractions test originally administered in the 1970s in England as part of the 'Concepts in Secondary Mathematics and Science' (CSMS) research programme documenting lower secondary students' understanding of mathematics. We present data from a recent study conducted in Denmark with 336 students aged between 12 and 14 years, presenting descriptive results as in the original research. In addition, we use Rasch modelling — a technique that was not widely used in the 1970s — to further analyse the students' understanding of fractions and to validate the test. Our findings indicate that Danish students' fractional understanding in word or diagrammatic contexts is slightly stronger in comparison to that of the English students in the original study, whereas the case for computations involving fractions is the reverse. We discuss the possible reasons for these differences. Our study provides evidence that fractional knowledge involves a number of components and suggest ways in which the test might be improved.

The Impact Sheet to this article can be accessed at 10.6084/m9.figshare.24219673

Published with license by Koninklijke Brill NV | DOI:10.1163/26670127-BJA10014

**Keywords**

conceptual replication – fractions – Rasch analysis

## 1 Introduction

In the 1970s, the English research program 'Concepts in Secondary Mathematics and Science' (CSMS) examined students' understanding of mathematics in 10 subjects in secondary mathematics (Hart, 1981). This was done through the development of diagnostic test items, which were piloted, refined, and validated through interviews with students who worked on assignments as well as through written tests. Student responses were coded to develop hierarchies across the investigated mathematical subjects; however, within each subject, certain levels of understanding were developed so that teachers could use the diagnostic tests as an evaluation tool in their teaching. Fractions was one of the subjects investigated; within this topic, four tests were developed: two age-graded tests with fraction problems involving word or diagrammatic contexts, each with a parallel test focusing on purely computational aspects with the same numeric items. Influenced by Piagetian theory (e.g., Shayer et al., 1976) and in order to communicate the results in a meaningful manner to teachers, policymakers, and curriculum developers, the test was designed with the aim of differentiating the 'levels' of conceptual understanding regarding fractions. Consequently, for each of the two problem tests, four hierarchical levels of understanding were identified (Hart et al., 1985).

Over the last four decades, the CSMS tests have been widely cited as providing evidence of the development of students' understanding in mathematics. However, it is important to recognise that the 'levels' are likely to have been influenced by wider societal and curricular factors as well as developmental factors. Recent work on learning trajectories suggests that development in fractions and rational numbers is not a simple hierarchy, but rather consists of several inter-related strands, or progressions, of key ideas (e.g., Confrey et al., 2009). Hence, in order to use these tests today, as a tool for teachers in their formative assessment of students' knowledge of fractions, we find it interesting to examine whether the levels appear to be of a hierarchical structure when analysing different data than the original. This not to disregard the idea of levels as a means of communicating overall understanding to teachers but to explore whether the existing levels are applicable to data collected in another context. Therefore, in this paper, we perform a conceptual replication (Schmidt, 2009)

of the Fractions test. Our focus is on CSMS Fractions 1 (Hart et al., 1985), the test targeted at younger students, and we include both the problem and the parallel computation item sets.

As indicated by Brown and Wood (2018), a replication study may include an extension to help clarify the interpretations of the original findings. In this paper, we report on a 'scaling out' extension (Melhuish, 2018), where we replicate the study in a different population (Danish rather than English) and a different time (2023 rather than 1976), thereby enabling the examination of whether the original study holds true beyond the context where it was developed (Aguilar, 2020) — for example, guarding against context-specific results (Schoenfeld, 2018). The second author conducted a partial replication of the CSMS tests in 2008 and 2009 using a restricted subset of the fractions items (Brown et al., 2010). However, we are only aware of one published full conceptual replication of the CSMS study which was conducted in Taiwan over three decades ago (Lin, 1989). This study found that the hierarchy of levels was largely consistent for Taiwanese students, but that range of understanding was narrower compared to the English students; moreover, contrary to the students in England, Taiwanese students performed better on computation items than on the other items.

Another extension in the current study is the use of a statistical tool that was not well-developed at the time of the original study — that is, using Rasch analysis to validate the measurement functioning of the Fractions test. Rasch adopts a similar but more sophisticated approach to the Guttman scaling approach used in the original study.[1] We do not have access to the original raw data and, therefore, we cannot know if the problems identified using the Rasch model would have been the same at the time of the original study. However, the analysis can add valuable knowledge to the usefulness of the test as a diagnostic tool today with regard to the levels of understanding originally identified. To facilitate comparison with the original study, and to allow a more meaningful interpretation by a general mathematics education audience, our findings are largely presented using item facilities (percentage correct) rather than the less intuitive Rasch measures. Finally, since the extension is performed in another country in a subsequent decade, where decimal notation is far more

1 Guttman scaling, like Rasch, assumes a hierarchical scaling of items and models the data as a unidimensional scale. See Hart and Johnson (1984) for additional information on the Guttman scaling model used in the original study. See the Supplementary Materials (available from the authors upon request) for further details on Rasch modelling and see Cascella et al. (2023) for further information on the use of Rasch modelling in replication studies.

prevalent within and outside classrooms than in England during 1976, we will need to examine any curricular changes that could have affected the changes in item facilities.

Therefore, the aim of this study is to answer the following four research questions:

1. To what extent does this test (still) constitute a valid assessment of students' knowledge of fractions?
2. In what ways can the test provide useful formative information to Danish teachers, a different context to that of the original study?
3. What are the strengths and weaknesses of Danish students' understanding of fractions and how does this compare to students from England over 40 years ago?

4) To what extent are the original CSMS findings (still) valid?

In this paper, we first summarise findings regarding fractions from the programme 'Concepts in Secondary Mathematics and Science' (CSMS). Next, we provide a brief description of changes in the curriculum during the last 50 years in both England and Denmark. Thereafter, we analyse data from the current study by 1) performing a Rasch analysis to examine whether the test constitutes a valid assessment, 2) describing the Danish results, and 3) comparing item facilities from the original study to those of the current study.

### 1.1 *Summarising Earlier Findings*

In CSMS, each test paper[2] presented problems (P notation) and a set of computations (C notation) designed to mirror the problems (Hart et al., 1981). The problem paper involved tasks posed either in words or using diagrams and intended to reveal students' conceptual understandings, whereas the computation paper aimed to reveal student's procedural knowledge using decontextualised tasks using the same calculations. In CSMS, samples were carried out in 1975, 1976, and 1977 (Hart & Johnson, 1980). Fractions 1 was only used in 1976; in that study, it was answered by 246 Year 1 students and 309 Year 2 students aged 11–12 and 12–13, respectively. Schools participated on a voluntary basis. Typically, recognition of a fractional name given to a region (e.g., a shaded section of a rectangle divided into equal parts) were correctly reproduced by most children in the study. However, addition and multiplication were much more difficult. Often, the same sample performed better on the word problems than on computations, thereby "suggesting that the methods used by the children were different for the two contexts" (Hart et al., 1989, p. 46).

2 The Tests and Marking Key can be found here: CSMS files (iccams-maths.org).

Across all subjects investigated in CSMS, an examination of all 'easy' items (i.e. those with high facilities) suggests that numerous secondary students worked entirely within the set of whole numbers (Hart et al., 1989). The results of the CSMS investigation revealed that the majority of secondary school children avoid using fractions, cannot generalize about them, and probably do not see them as an extension of the set of whole numbers (Hart, 1981). When fraction computations involved addition and subtraction, the percentage of first year students succeeding was always higher than students in every other year (Hart et al., 1981).

In CSMS, four levels of fractional understanding were identified. The levels were hierarchical in order, thereby implying that a student attaining level three would most likely also attain levels two and level one. At level one, students could make meaning of simple fractions ($\frac{1}{2}$, $\frac{1}{5}$, $\frac{2}{3}$) by enumerating pieces of a whole in a simple context or diagram. At level two, the students would further understand the meaning of a fraction using discrete quantities, obtain equivalent fractions by doubling, and perform addition of two fractions with the same denominator. At level three, students could also use equivalent fractions to name parts, use equivalence with less familiar fractions, and order unit fractions. At level four, students would be able to use their knowledge of fractions when more than one mathematical operation was required (Hart et al., 1985). Other aspects of rational numbers were assessed using different tests, notably tests of ratio, place value, and decimals and measurement.

### 1.2 *Difficulties with Fractions Faced by Students*

In CSMS, the authors concluded that many children use their own informal methods instead of formally taught methods. However, the findings of Kerslake (1986) indicate that it is different when examining fractions where "children are seen to rely on rote memory of previously learned techniques" (Kerslake, 1986, p. 87), which often results in half-remembered rules being inappropriately applied. These findings are supported by studies from the US that reveal that middle-level students rely on their rote memory of rules to solve fraction problems (Kieren, 1988). Fraction instruction is often described by teaching where students routinely learn how to perform fractions operations every year and then forget how to perform them (Aksu, 1997). This kind of instruction, as well as the structure of fractions, is revealed to be the basic reason for difficulties in learning fractions and rational numbers (Streefland, 1991). This is despite the fact that content knowledge in a complex domain such as rational numbers and fractions do not depend on target concepts and operations (Lamon, 2007) since they require reorganisation of knowledge of

whole numbers and operations on these numbers (Carraher, 1996). Moreover, concepts and operations represented by children's natural language are used in their construction of knowledge of fractions (Steffe & Olive, 1990).

Generally, children are familiar with part of a whole model of fractions (Kerslake, 1986), which is identified as the most common starting point in fraction instruction (Baturo, 2004). This is also the case in Danish textbooks where 60%–70% of the tasks involve counting activities, labelling of fractions, and different parts of whole activities (Færch & Pedersen, 2023). This is despite the fact that the part of a whole interpretation is considered to "inhibit the development of other interpretations of a fraction" (Kerslake, 1986, p. 89). However, the part of a whole interpretation of fractions is just one of five interpretations of rational numbers widely accepted in the research literature: part-whole, measure, operator, ratio, and quotient (Kieren, 1993; Behr et al., 1983; Lamon, 1999). To understand rational numbers, students need experiences with different interpretations — not only as objects of computation (Kieren, 1976). However, not all interpretations of fractions "provide equal access to deep understanding and no single interpretation is a panacea" (Lamon, 2012).

Whole number schemes can interfere with students' efforts to learn fractions (Behr et al., 1984) often by students processing numerator and denominator as two separate whole numbers (Pitkethly & Hunting, 1996), thereby discarding $\frac{a}{b}$ as a number (Hannula, 2003). Carraher (1996) refers to this as the cardinal sin, thereby indicating that counting and matching activities in fraction instruction with an emphasis on part of a whole interpretation of rational numbers causes students to focus on the cardinal number. Students who see numerator and denominator as two separate numbers are seen to inappropriately apply natural number properties to the concept of rational numbers, which is occasionally referred to as natural number bias (Ni & Zhou, 2005). The natural number bias can cause trouble with the ordering of fractions (e.g., $\frac{1}{5}>\frac{1}{3}$) and addition and subtraction of fractions, where students add and subtract numerators and denominators as if they were natural numbers (e.g., $\frac{1}{4}+\frac{1}{2}=\frac{2}{6}$) (Gabriel et al., 2013; Nunes & Bryant, 1996). Further, students are seen to perform well in addition and subtraction of fractions with the same denominator, while performance decreases dramatically when the denominators are different; moreover, many errors in computation with fractions are linked to additive reasoning (Gabriel et al., 2013).

In Denmark, little research has been conducted on knowledge regarding how students understand, and misunderstand, fractions or rational numbers. One study from 2022 examines differences in fraction learning among high- and low-performing Danish 11–12-year-old students (Pedersen et al., 2022). The authors conclude that while high-performing students learn

fractions continuously throughout the observed school year, low-performing students only learn fractions when instructed directly in fractions. The article was part of a PhD project examining 11–12-year-olds difficulties with the concept of fractions. The only other published study in a Danish context investigates pre-service teachers' knowledge of operations on rational numbers (Putra, 2018). In that study, pre-service teachers collaborate to solve four tasks: addition and subtraction of fractions and multiplication and division of decimals. The study concludes that Danish pre-service teachers prefer to teach fractions in real-life situations. The context of real-life situations, particularly their use of a pizza-representation of fractions, is doable for the pre-service teachers when teaching addition and subtraction tasks; however, when faced with multiplication and division tasks, their preferred representation makes it difficult to continue in the real-life situations.

## 2 Background

### 2.1 *Development of Fraction Instruction in England Since the 1970s*

The last four decades have seen significant changes in mathematics education in England (see Hodgen et al. (2022) for a discussion). In 1976, outside the mathematics classroom, imperial measures were still widespread, which implies little use of decimal notation and almost no use of calculators. Because there was no national curriculum at the time in England, the school curriculum was largely structured by the syllabi of a variety of examination boards. The school-leaving age was raised to age 16 in 1972 and, in 1976, numerous students left school at the age of 16 with little or no formal qualifications. Partially influenced by the CSMS study, the National Curriculum was introduced in 1989 with the aim of raising attainment in mathematics (Brown, 1996). Since then, several revisions have been made to the National Curriculum. In addition, the national testing was introduced at ages 7, 11, and 14 between 1991 and 1995,[3] along with greater regulation and accountability measures (Millett & Johnson, 2000). In 1998, the National Numeracy Strategy was introduced into primary schools and subsequently into lower secondary schools (2001), thereby placing greater emphasis on calculation, particularly mental calculation (Brown et al., 2000). One of the more significant changes for the purpose of this paper is that, over time, greater emphasis has been placed on measurement and computation with decimals rather than with fractions. Moreover, significant importance has been placed on calculation, both mental and paper-and-pencil

3 National testing at 14 was later abolished in 2008.

based and while the use of calculators is permitted, this has been discouraged in primary schools (Hodgen, 2012).[4]

### 2.2 *Development of Fraction Instruction in Denmark Since the 1970s*

In Denmark, primary and lower secondary school is called 'Grundskolen' and is for students aged 6 to 16. These 10 years include a transition year called Kindergarten class. The remaining nine years of compulsory school are subject to the same legislation with the overall goal of covering three consecutive school years. There have been several changes in the legislations regarding mathematics in primary and lower secondary school in Denmark since the 1970s. However, goals for fraction instruction have only been subject to minor changes (Haahr & Jensen, 2008). Three legislations are worth mentioning in relation to the current study. In 'Skoleloven af 1958' (Undervisningsministeriet, 1960), fractions were mentioned in years 4–7 with a separate curriculum for each year. In 'Lov om folkeskolen' (Helsted, 1975) from 1973 to 1993, this was no longer the case. Currently, there are overall goals for years 4–6 and 7–9, respectively, with the mention of fractions. Years 4–6 now boils down to, "At these grade levels, calculations with simple fractional numbers includes addition and subtraction, and very simple examples of multiplication and division of fractions. It must be considered inappropriately, to work with division, where the divisor is given as a fraction or a decimal number" (Helsted, 1975, own translation).

However, the subject of mathematics in Denmark cannot be seen solely in the light of the official curriculum and teaching guidelines, as both textbooks and final exams have been decisive for the teaching of the subject in Denmark since the 1950s (Haahr & Jensen, 2008). The following paragraph is based on two textbook systems from the 1970s, identified in Haahr and Jensen (2008), and two of the most used in the 2020s, identified in Færch and Pedersen (2023). The textbooks from the 1970s include years 4–7 from Matematik (Cort & Johannessen, 1970, 1973, 1974, 1975) and Matematiktimen (Clausen & Jensen, 1971, 1975, 1976, 1977), and the newer ones include Kontext+ (Andersen et al., 2014, 2015, 2016; Hansen et al., 2015, Jensen et al., 2016), and Multi (Bisgaard et al, 2015; Kristensen & Teglskov, 2012; Lathi et al., 2016; Mogensen et al., 2014, 2015).

In the textbooks from both the 1970s and the 2020s, many different fraction models have been used, including circle, area, number lines and discrete quantities. At the same time, all textbooks treat fractions in an everyday context,

4 Since the 2014 revision of the National Curriculum, more emphasis has been placed on fractions and formal written calculation methods and less emphasis is placed on mental calculation.

including word problems, where examples from the students' everyday life are focused upon. Moreover, *Matematik* also utilises of block models. There are probably greater differences among various textbook systems than can be found between the two time periods, when only considering how the concept of fractions is treated. However, the focus on fractions is treated in greater depth in the textbooks from the 1970s, thereby implying that the concept of fractions is treated several times a year over four school years and covers a larger proportion of the textbook's pages; in contrast, in newer books, they are only treated in a single chapter for each of the four years. It also appears that students from the 1970s generally encountered a larger variety of fractions, including examples of what a fraction was not. In comparison, fractions in textbooks from the 2020s mainly focuses on proper fractions and subsequently on improper and mixed fractions. However, there are no examples of fractions as $\frac{2}{2}$, $\frac{0}{3}$, and only a few examples of fractions written as $\frac{12}{3}$, which is the case in Cort and Johannessen (1970), where examples of what a fraction is not are also included. On the other hand, textbooks from the 2020s focus more on students' own productions of fractions, shifts in representations of fractions, and use of concrete materials like centicubes. In both decades, equivalence is mainly dealt with as procedures used to reduce and expand fractions, where students are asked to find one or more equivalent fraction(s), sort fractions, and identify the smallest/largest fraction. None of the textbooks require students to find a denominator or numerator that will produce one fraction that is equivalent to another fraction. In all textbooks, computation with fractions is treated through the introduction of an algorithm supported by different visual areal models of the procedure, but these algorithms are introduced earlier in the textbooks from the 2020s, thereby focusing less on understanding the concept of fractions. When examining the textbooks, there does not appear to be any major differences in the introduction of fractions nor in the connection between fraction and decimals, except that the transition from fractions to decimals has changed from algorithms that use long division to the use of a calculator.

## 3 Methodology

### 3.1 *Description of Danish Participants*

In March and April 2023, the CSMS Fractions 1 and the related Computation was given to 336 Danish students distributed on 17 classes at six different schools. The tests were performed by year 6 and 7, covering students aged 12–13 and 13–14, respectively. Ending up with test data from 107 year 6 and 229 year 7.

The schools were not randomly chosen but had volunteered to participate in the trial. All participating schools are in a radius of 65 kilometres from the Danish capital, Copenhagen. However, school data from Uddannelsesstatistik .dk (Educational statistics in Denmark) shows that the participating schools are widely distributed in both socioeconomic backgrounds and in average grades looking at the compulsory exams after grade 9 in Denmark. In most Danish classes there are dyslexic students and bilingual students which was also the case for the students involved in this study. In some of the classes, the dyslexic students had the possibility to have the test read out load via their computer. In other classes, the students did not wish to stand out and therefore refused to get help reading the tasks. A few newcomers (immigrants) had the tasks translated into English.

### 3.2 *Implementation of the Fractions 1 Test*

The test comprises two parts: The first part, Fractions 1, comprises 42 items involving problems using words or diagrams; the second part is the computation with 18 purely mathematical calculation items, most of which corresponding to a problem item. In the Danish study, the tests were conducted based on the original CSMS Teacher's Guide (Hart et al., 1985) and the students were informed that

- they could get help understanding or reading the tasks;
- they would not get help answering the tasks;
- if they did not know how to solve the task, they should move on to the next task;
- they only needed a pen and if they wished to change something, they should cross out the old answer; and
- they should bring a book or some homework to occupy them in case they finish early.

Most classes worked for 30–40 minutes twice with a small break in between. While a few students finished after 20 minutes, others needed additional time. Some of the students finished the next day, while others either did not have the opportunity or the desire to do so. All students were told that the questions were not a test, but a means for us to explore how Danish students understand fractions. This was important to some of the teachers since they had students that would panic when hearing the word test or would stay back at home if they knew about it in advance. Simultaneously, it was a way to tell the students about the importance of having their individual answers and not copy answers from their peers since collaborative work is rather common in Denmark whereas testing is not.

The test was translated into Danish by the first author of this paper and was piloted with one Danish student prior to the full administration. In this pilot, one word in item number P18 was difficult to understand: 'bicycle spokes'. The wording was not changed, but an introduction to the word was instead made part of the introduction to the test. Yet, that task was one of the tasks that students had the most questions about before engaging with it. However, we do not know whether this has something to do with the translation or simply because Danish students do not know the meaning of the word, unless they are cycle enthusiasts themselves or have one in the family. This is despite the fact that 93% of 6–14-year-olds in Denmark own a bike (DTU, 2021).

Contrary to the original study, all students were not given the tasks in the same order. This was done to avoid collaborative work. In the first two schools, half the students began with problem tasks and the other half with computation tasks. Unfortunately, the computational tasks appeared rather demotivating, particularly for the lower performing students. Therefore, the second half of the schools were handed out tests, where half began with question P1 and the other half with question P10, all ending with the computation tasks.

### 3.3 *Marking of the Fractions 1 Test*

When marking both tests, the *Fractions 1: Marking Key from the Teacher's Guide* (Hart et al., 1985) was used. In the marking key, code 1 was used for correct answers, codes 2–8 for interesting or typical errors, code 9 for other incorrect answers, and code 0 for missing responses. The same person marked all the tests to ensure that the marking was done in the same manner. Whenever the same response came up repeatedly, a new code was added to the marking key, and all already marked tests were looked at again to see if that response was overlooked in these. New codes were also added if an interesting answer appeared — for example, if the wrong answer was due to an overgeneralisation of addition with whole numbers. However, these codes were ultimately deleted at the end if only one to three student answers fell into the coding category. New codes are marked with yellow in Appendix 1. In 'Fractions 1: Computation', error codes where not part of the marking key (Hart et al., 1985) and, thus, error codes were designed, first, by examining corresponding tasks from 'Fractions 1: Problems'. These codes are marked blue in Appendix 1. Second, the error codes were designed by examining typical student answers to the tasks. The first type of error codes was retained regardless of how many student answers were identified, whereas the second type of error codes were deleted again if used in less than 3% of the total number of responses. Responses for these 17 codes where then changed to code 9 instead.

To compare the item facilities from the current study with item facilities from the original CSMS Study, some of the original data where reinterpreted since the original Marking Scheme (Hart & Johnson, 1980) does not correspond to the *Marking Key from the Teacher's Guide* used in this study (Hart et al., 1985). All percentages corresponding to an error code from Hart and Johnson (1980) where moved to match the error codes from Hart et al. (1985).

### 3.4 *Statistical Methods*

Descriptive statistics were initially produced using Excel, whilst the Rasch analysis was conducted in R using the eRm package (Mair et al., 2021) and then replicated in jamovi (The jamovi Project, 2022) with only minor differences in item fit statistics between the two programmes. All missing responses were treated as incorrect, thereby implying that for the Rasch analysis, the test comprises 59 dichotomously scored (right/wrong) items on fractions.

In this paper, the goal with the Rasch analysis is not to improve the test for future use but to assess whether the items form a unidimensional scale such that the test provides a comparative measure of student understandings of fractions. To do this, we identify items that threaten the integrity and, thus, the validity of the existing test. Briefly, the Rasch model is a probabilistic model based on an item response theory (IRT) model (Hambleton, 1993). It can be used to estimate person 'abilities'[5] and item difficulties. It assumes that these latent variables, person ability and item difficulty, can be measured on the same unidimensional interval scale. Furthermore, it assumes local independence of items — for example, performance of pairs of items should be independent. Our analysis focuses on items that reveal some misfit from the model and we focus on two measures of item fit: infit and outfit. Infit is more sensitive to responses close to person abilities, whilst outfit is more sensitive to outlier responses far from person abilities. Both statistics have an expected value of 1. Values above 1 indicate unpredictability, while values below 1 indicate redundancy. The range of acceptable values for item fit depend on the purpose of the test and narrower ranges used for tests with higher stakes (Bond & Fox, 2007). Because this is a low stakes test with a diagnostic purpose, we adopt Linacre's (2002) guidance of item fit statistics in the range of 0.5–1.5 as being productive for measurement. Values >2.0 are judged as potentially distorting and, for a sufficiently valid test, only a small number of items should have such values.

5 Following the Rasch tradition, we use the term 'ability' for the latent trait of students' understanding of fractions as measured by the test, and not to indicate a general notion of mathematical ability.

We note that in the original CSMS research, 14 items were judged not to fit the model sufficiently well. These items were excluded from the hierarchy but were included in the test because they were nevertheless considered to provide useful information regarding students' understanding.

## 4 Results

In this section, we first report the results of the Rasch analysis to validate the Fractions 1 test and examine whether the hierarchical levels of understanding initially identified are still applicable to the current data. Then, we present descriptive results of the Danish participants and compare these with the English results from 1976.

### 4.1 *The Rasch Analysis*

Overall, the 336 students' test scores ranged from 0 (no correct items) to 57, thereby implying that no student achieved the maximum of 59 points. The average score was 29.7 (or 50% of the maximum score), with a standard error of 12.8 (see 13).

The Wright Map (Figure 1) presents the distribution of person abilities and item difficulties. We remind the reader that, in Rasch modelling, these latent variables, person ability, in this case understanding of fractions, (on the left of the diagram) and item difficulty (on the right of the diagram), can be measured on the same unidimensional interval scale. The Rasch scale is shown on both the right and the left of the diagram.[6]

Broadly, the Wright map shows a balanced distribution of person abilities. As might be expected, given the original study's finding of a hierarchy of understanding (Figure 1), the distribution of item difficulties shows some clustering. At the bottom of the map, there are five rather easy items involving counting and labelling tasks. What makes these particularly easy is that all

TABLE 1 Summary of test scores across the Danish sample

| *N* | Minimum | Maximum | Mean | Median | SD | SE | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| 336 | 0.0 | 57.0 | 29.7 | 30.0 | 12.8 | 0.70 | −0.12 | 2.20 |

6 Technically, this scale is measured in logits. See Cascella et al. (2023) for further information.

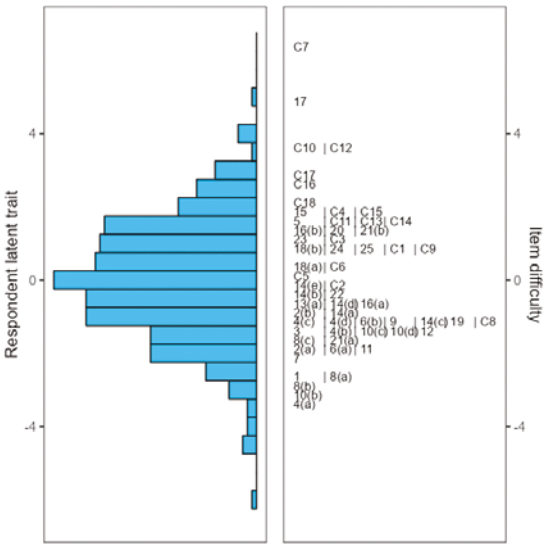


FIGURE 1
Wright map depicting distribution of person estimates and item parameters

tasks can be solved correctly using only knowledge of whole numbers to count parts and wholes to label a fraction or colour fractional parts only by looking at the whole number in the numerator. In the second group of items, there are items that involve the labelling of fractions using discrete units, items focusing on the size of a fraction, equivalence, and addition of fractions. The third group includes items with more than one operation, often items for which finding equivalent fractions are necessary to compute the answer. The last group includes computational tasks with either division of fractions or subtraction with mixed numbers.

Overall, the analysis indicated a reasonably good fit to the Rasch model with all the infit values falling within the productive range. Nine items had outfit values beyond the 0.5–1.5 productive range (Table 2).[7] Of these, only three have potentially distorting values, all close to the 2.0 threshold, and, hence, are judged not to be a threat to the validity of the test.

In Table 2, items with no indication of CSMS levels were not part of the levels of understanding in the original project.

Overall, the 2023 Danish administration suggests that the test performs at least as well as in the original study.

7 Full details of all item facilities, Rasch measures, and infit and outfit values are available from the authors upon request.

TABLE 2 Nine items with outfit values beyond the productive range

| Item | Item description | CSMS level | Facility | Infit | Outfit |
|---|---|---|---|---|---|
| P4a | Shade two-thirds of three | 1 | 0.93 | 0.89 | 2.09 |
| P8c | What fraction is shaded? (Triangle showing seven-sixteenths) | | 0.75 | 1.41 | 2.07 |
| P5 | What is 17 cm divided into four equal pieces? ($4\frac{1}{4}$) | | 0.26 | 1.30 | 2.02 |
| P10b | Mark the greater fraction ($\frac{3}{7}$ or $\frac{5}{7}$) | | 0.90 | 1.08 | 1.60 |
| P15 | Divide three bars of chocolate equally between five children? ($\frac{3}{5}$) | | 0.22 | 1.20 | 1.57 |
| P8a | What fraction is shaded? (showing four of eight in a rectangle) | 2 | 0.86 | 1.09 | 1.56 |
| P10d | Mark the greater fraction ($\frac{1}{5}$ or $\frac{3}{10}$) | | 0.71 | 1.35 | 1.54 |
| P4d | Shade two-thirds of nine. | 3 | 0.70 | 1.04 | 1.51 |
| C17 | $32\frac{2}{3}+5\frac{1}{4}$ | | 0.11 | 0.82 | 0.49 |

### 4.2 *Hierarchy Levels in CSMS*

The levels of understanding in CSMS was obtained by correctly answering $\frac{2}{3}$, $\frac{7}{10}$, $\frac{7}{11}$, and $\frac{3}{4}$ of Level 1, 2, 3, and 4, items respectively (see Table 3 for description of items) (Hart et al., 1985).

In Figure 2, the proportion of Danish students at each level of understanding are compared to data for the 13-year-old English students (Hart et al., 1981), thereby indicating only minor differences between the two countries when examining only the CSMS levels of understanding.

The item facilities (Table 3) for level 2 and level 3 does not clearly identify level 3 items as more difficult than level 2 items. Even though there are only a few students who obtain levels 2, 3, or 4 without also obtaining the lower levels, the item facilities for the Danish data reveals that there are quite a few level 2 items with lower facilities than the items at level 3 (see Table 3). This is mostly evident for level 2 and level 3 items. When adding a label that includes the mathematical idea that students are expected to engage with when solving

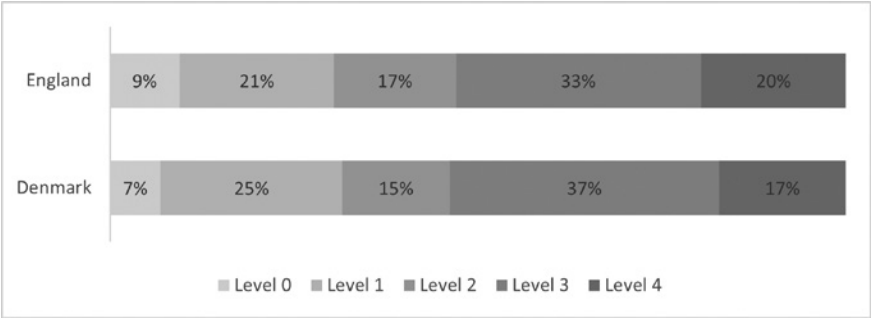


FIGURE 2 Bar chart representing the proportion of students at various hierarchy levels in Fractions 1

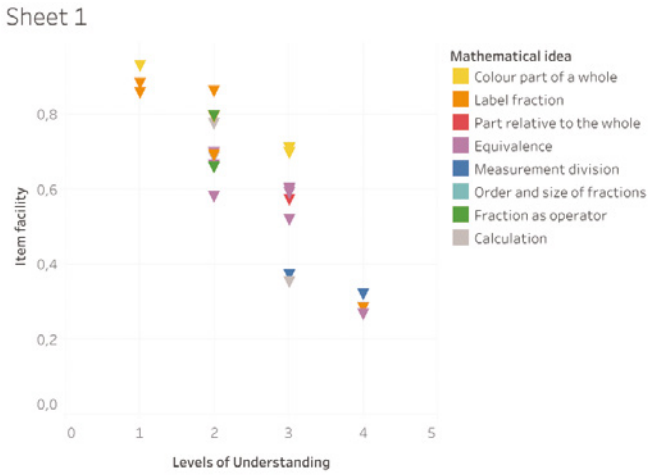


FIGURE 3 The distribution of item facilities according to the four CSMS levels of understanding

the task (Figure 3), it is evident that level 2 and level 3, particularly, are a mixture of numerous different mathematical ideas.

### 4.3 *Strengths and Weaknesses of the Danish Students*

Figure 4 presents the distribution of the Danish students according to the test score.

In the following section, the Danish results will be accounted for by closely examining the students' strengths and weaknesses according to the Fraction 1 test.

Generally, the item facilities are high (above 60%) for items that involve colouring a figure representing a fractional part, labelling of fractions, the size

TABLE 3 Item description and facility for each test item identifying the levels of understanding

| Item | Item description | CSMS level | Facility (Denmark) |
|---|---|---|---|
| P4a | Shade two-thirds of three. | 1 | 92.6 |
| P8b | What fraction is shaded? (Showing four of eight in a circle) | 1 | 88.1 |
| P1 | Share a stick of rock equally among 5 children. | 1 | 85.7 |
| P8a | What fraction is shaded? (Showing four of eight in a rectangle) | 2 | 86.0 |
| P2a | Put a ring around $\frac{1}{3}$ of 12 marbles. | 2 | 79.5 |
| P6a | What fraction is cracked when 5 of 12 eggs are cracked? | 2 | 79.2 |
| P21a | What fraction does John have left if he pays $\frac{3}{5}$ in tax? | 2 | 77.4 |
| P14c | $\frac{5}{10}=\frac{\square}{30}(15)$ | 2 | 69.6 |
| P6b | What fraction is not cracked when 5 of 12 eggs are cracked? | 2 | 69.0 |
| P19 | What fraction of the flour has been used when both $\frac{3}{8}$ and $\frac{2}{8}$ has been used? | 2 | 68.5 |
| P14a | $\frac{1}{3}=\frac{2}{\square}(6)$ | 2 | 66.4 |
| P2b | How many marbles is $\frac{2}{3}$ of 12 marbles? | 2 | 65.5 |
| P14b | $\frac{6}{8}=\frac{3}{\square}(4)$ | 2 | 58.0 |
| P4b | Shade two-thirds of six. | 3 | 70.8 |
| P12 | Put the fractions in decreasing order. (Four unit fractions) | 3 | 70.8 |
| P4c | Shade two-thirds of six. | 3 | 69.6 |
| P4d | Shade two-thirds of nine. | 3 | 69.6 |
| P16a | $\frac{2}{7}=\frac{\square}{14}(4)$ | 3 | 60.1 |
| P14d | $\frac{2}{3}=\frac{\square}{15}(10)$ | 3 | 59.8 |
| P13a | Who eats most if Peter eats 4 of 8 pieces and Abdul eats 2 of 4 pieces? (The whole being equally big) | 3 | 58.9 |
| P22 | What fraction of the whole disc is shaded when one-sixth of three-fourths is shaded? | 3 | 57.1 |
| P14e | $\frac{4}{12}=\frac{1}{\square}(3)$ | 3 | 51.8 |

TABLE 3 Item description and facility for each test item identifying the levels of understanding (*cont.*)

| Item | Item description | CSMS level | Facility (Denmark) |
|---|---|---|---|
| P24 | How many runners are needed to $\frac{3}{4}$ run km if each of them run $\frac{1}{8}$ km? | 3 | 36.9 |
| P18b | What length of wire is left after cutting three spokes of $10\frac{1}{2}$ cm? | 3 | 35.1 |
| P23 | How many pieces of wood $1\frac{1}{4}$ cm long can we get from a piece $8\frac{3}{4}$ cm long? | 4 | 31.8 |
| P21b | What fraction does John have left if he pays both $\frac{3}{5}$ in tax and $\frac{1}{10}$ on rent? | 4 | 28.6 |
| P20 | What fraction of the floor is tiled? (The tiled in twenty-fourths and the non-tiled in twelfths) | 4 | 28.0 |
| P16b | $\frac{2}{7}=\frac{x}{14}=\frac{10}{\square}(35)$ | 4 | 26.5 |

of a fraction, subtracting a proper fraction from one, and addition with common denominator. There are only minor differences in item facilities for P4b–d (Table 4). The difference in item facility comparing P4a to the other items in P4 indicates that approximately 20% of the Danish students only focuses on the numerator — colouring two instead of two-thirds corresponding to shading $\frac{1}{3}$ in P4b and P4c. In P4d, this strategy would result in colouring $\frac{2}{9}$, which is not coded. However, code 9 includes the colouring of $\frac{2}{9}$, thereby accounting for 10% of the responses, compared to 0.5% code 9 responses in P4a–c. Only few of the Danish students coloured half the figure.

Further, labelling of proper fractions, understood as a part of a whole relationship, also appears to be familiar to the Danish students, as indicated in Table 5. In P1, half of the incorrect responses are from students who marked the division on the sketch, with some more accurate than others. Other incorrect answers included students measuring the stick using a ruler to write how many centimetres each child would get. The easiest labelling task is P8b in which the students are to identify the fraction $\frac{4}{8}$. While some students write $\frac{4}{8}$ or $\frac{4}{8}=\frac{1}{2}$, many students answer with $\frac{1}{2}$, thereby indicating that some students do not count the parts when the fraction is a representation of one half. In P8c, 7% of the incorrect responses comprises students who did a part-part comparison, thereby yielding the answer $\frac{7}{9}$. Part-part comparisons only apply for

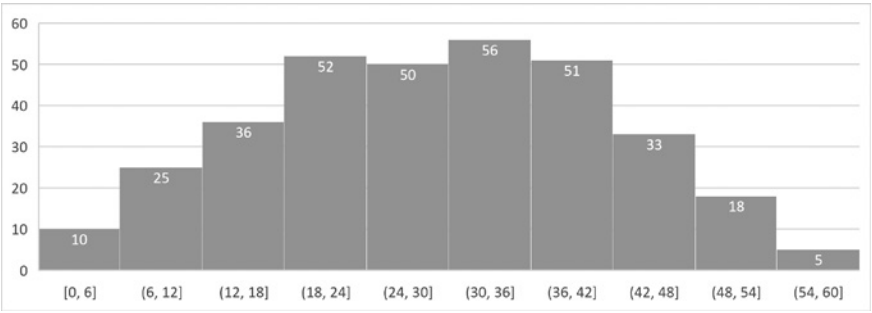


FIGURE 4 Distribution of Danish students by test score

TABLE 4 P4. Shade two-thirds of the shape

| | 4A | 4B | 4C | 4D |
|---|---|---|---|---|
| |  |  |  |  |
| Correctly Shade $\frac{2}{3}$ | 93% | 71% | 79% | 70% |
| Shade $\frac{1}{3}$ | 1% | 5% | 4% | 0.3% |
| Shade $\frac{2}{9}$ | 1% | 17% | 19% | 13% |

TABLE 5 Labelling of fractions

| Item | Item description | Correct | Incorrect | Missing |
|---|---|---|---|---|
| P1 | Share a stick of rock equally among 5 children. | 86% | 9% | 5% |
| P6a | What fraction is cracked when 5 of 12 eggs are cracked? | 79% | 8% | 13% |
| P8a | What fraction is shaded? (Showing four of eight in a rectangle) | 86% | 7% | 7% |
| P8b | What fraction is shaded? (Showing four of eight in a circle) | 88% | 5% | 7% |
| P8c | What fraction is shaded? (Triangle showing seven-sixteenths) | 75% | 17% | 8% |

3.6% and 1.2% of the answers in tasks P8a and P8b, respectively. The remaining incorrect responses is a mixture of students committing errors in counting the parts and, thus, ending up with answers like $\frac{7}{15}$, $\frac{7}{14}$, or $\frac{8}{16}$, or also guessing $\frac{1}{2}$ by merely looking at the figure.

In the test, there are several problem tasks for fraction equivalence. One of these types of tasks asks students to identify the greater fraction among two fractions, as seen in Table 6.

However, the decrease in facilities from the first item indicates that at least 20% of the students compare fractions using other strategies — for example, whole number strategies. In the first task, students can get the correct answer even though they only compared the whole numbers in the numerator and ignored the denominator. This strategy also led to the correct answer in the third task. Since the numerators are the same in the second task, another strategy is required. One whole number strategy is finding the difference between the numerator and the denominator, where selecting the larger difference will lead to an incorrect answer in the second but a correct answer in the third. The two different whole number strategies yielding the correct answer in the third task makes it difficult to identify if the students compared fractions or whole numbers. If a student coded Correct — Incorrect — Correct, this can indicate that the students are either guessing or using whole number strategies — 15% of the students answers are coded in this manner in the current study. Another 16% coded Correct — Correct — Incorrect, thereby indicating that these students are guessing the answers. However, the item facility when students are to order four unit fractions (P12, Table 3), beginning with the smallest, is more or less the same (e.g., 71%); this indicates that the students who does get the second and third tasks correct utilises a strategy either comparing fractions as numbers or at least knows the rule that the bigger the denominator, the smaller the fraction, thereby making the ordering of unit fractions very easy. However, 17% of the Danish students order the fractions only by looking at the denominator, reversing the order of fractions.

Danish students appear to have grasped the idea that the parts taken together must exhaust the whole (see Table 7), with items P21a and P6b being the items involving fraction computations that has the highest item facilities. In P6b, students can perform the computation with whole numbers before labelling the fraction. However, this strategy is not possible in P21a, which has a slightly higher facility than P6b, thereby indicating that students do, in fact, use knowledge of the whole as $1=\frac{b}{b}$, counting from a to b before labelling a new fraction $\frac{b-a}{b}$.

TABLE 6 P10B–D. Put a ring around the greater fraction

| | $\frac{5}{7} > \frac{3}{7}$ | $\frac{3}{4} > \frac{3}{5}$ | $\frac{3}{10} > \frac{1}{5}$ |
|---|---|---|---|
| Correct | 90% | 72% | 71% |
| Incorrect | 2% | 19% | 20% |
| Missing | 7% | 9% | 8% |

TABLE 7 Parts taken together must exhaust the whole

| Item | Item description | Correct | Incorrect | Missing |
|---|---|---|---|---|
| P21a | What fraction does John have left if he pays $\frac{3}{5}$ in tax? | 77% | 4% | 19% |
| P6b | What fraction is not cracked when 5 of 12 eggs are cracked? | 69% | 17% | 14% |
| C4 | $1-\frac{5}{12}$ | 22% | 27% | 52% |
| C15 | $1\frac{1}{5}-\frac{3}{5}$ | 22% | 19% | 59% |

The low facilities in the corresponding computation tasks, C4 and in a slightly more difficult version C15, indicated that the students do not solve the problem items by performing the computation $1-\frac{3}{5}$. From a computation aspect, it does not appear to make any difference to the Danish children if the minuend is a whole number or a mixed number, at least not when looking at the item facility. The types of errors in the problems and the computation have different characteristics. While the incorrect answers in P6b are related to either understanding the question, or difficulties labelling a fraction correctly leading certain students to invert the fraction, the incorrect in the computation items are related to misunderstood algorithms. The type of errors in C15 is related to students either ignoring the fraction part of the mixed number and finding the difference between 1 and the fraction $\frac{3}{5}$ or ignoring the commutative law and subtracting 1 from 3. Thus, 5.7% of the students give the answer $\frac{2}{5}$, while another 4.5% answer $1\frac{2}{5}$.

The only other computational aspect of fractions with item facilities above 60% is addition of fractions with the same denominator. There does not appear

TABLE 8 Fraction computation

| Item | Item description | Correct | $\frac{a}{b}+\frac{c}{d}=\frac{a+c}{b+d}$ | Other incorrect | Missing |
|---|---|---|---|---|---|
| P12 | What fraction of the flour has been used in total when $\frac{3}{8}$ and $\frac{2}{8}$ has been used? | 69% | 3% | 7% | 21% |
| C8 | $\frac{3}{8}+\frac{2}{8}$ | 68% | 7% | 5% | 20% |
| C9 | $\frac{1}{10}+\frac{3}{5}$ | 35% | 18% | 17% | 30% |
| C14 | $\frac{1}{3}+\frac{1}{4}$ | 25% | 17% | 18% | 40% |

to be any difference between problems and computation for the Danish students with regard to this aspect.

Students are quite comfortable adding fractions with a common denominator irrespective of the context (see Table 8). However, only few of them are able to use their knowledge of equivalent fractions to find a common denominator before performing the addition, thereby leading them to answer by adding denominator with denominator and numerator with numerator. This type of error is found in approximately 17% of the responses when the denominators are different (e.g., items C9 and C14, Table 8) compared to only in 3–7% of the responses when there is a common denominator (e.g. items P19 and C8; Table 8). It appears that the more complex the fraction addition becomes, the more students apply misremembered rules for computation.

Tasks that cause more trouble for the Danish students primarily involve computation tasks, particularly tasks involving division of fraction or mixed numbers, multiplication involving fractions, addition and subtraction of fractions with different denominators, and fractions as indicated division regardless of the fraction being a proper or improper fraction. Moreover, problem tasks where two or more computations are needed were very difficult for most students. These include a. tasks where students are required to find a common denominator before solving the task, b. tasks with fair sharing where the answer requires students to recognise a fraction as a number, and c. labelling of fractions when the denominator and the numerator are given as different units.

### 4.4 *Comparison of Item Facilities*

In the original study, Fractions 1 was administered to two separate year groups, thereby yielding separate data for the two groups. For the complete original

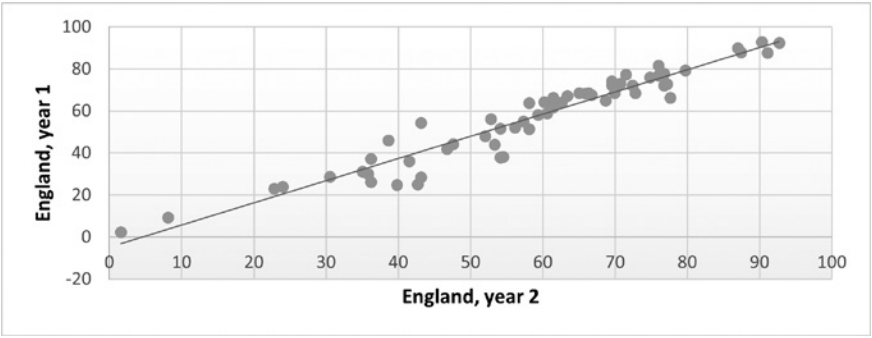


FIGURE 5 Scatterplot of 58 matched item facilities for English students aged 12–13 compared to English students aged 11–12 years (anno 1976)

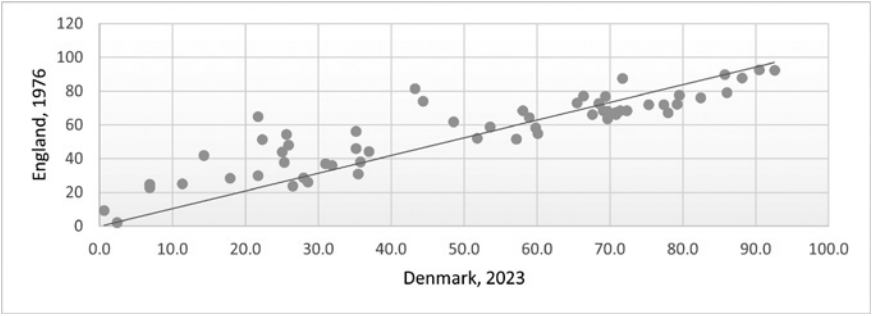


FIGURE 6 Scatterplot of 58 matched item facilities for Denmark anno 2023 compared to England anno 1976

data, we refer to Hart and Johnson (1984). For the sake of comparison, we only examine the results for the oldest students, those in year 2, since they are closer in age to the Danish students and since there are generally small differences between students in years 1 and 2, as depicted in Figure 5.

A graphic presentation of differences in item facility is depicted in Figure 6. Generally, the item facilities are lower in the current study now than they were in 1976. This is particularly the case for computation tasks. However, for a large proportion of the items, the item facility remains more or less the same (30%) or with a slight increase (20%).

It appears that there are a large number of items with only minor differences and a few items with an increase in facilities for the Danish students. However, the major change in facilities is due to decreases in the facilities in Danish students' results. The item facilities are generally lower for the Danish students, but when examining the problems and computations as two separate tests, it looks as though the main differences lie in the computation task.

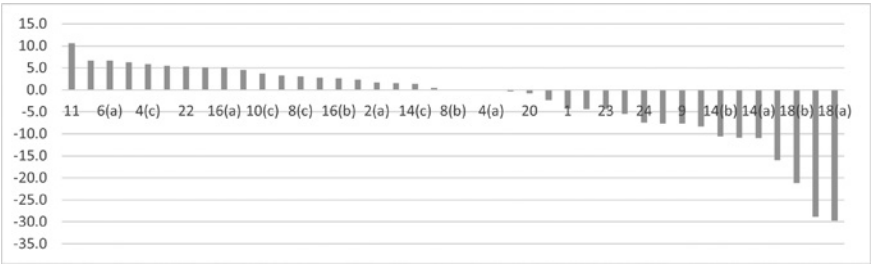


FIGURE 7 Change in item facility for problem tasks

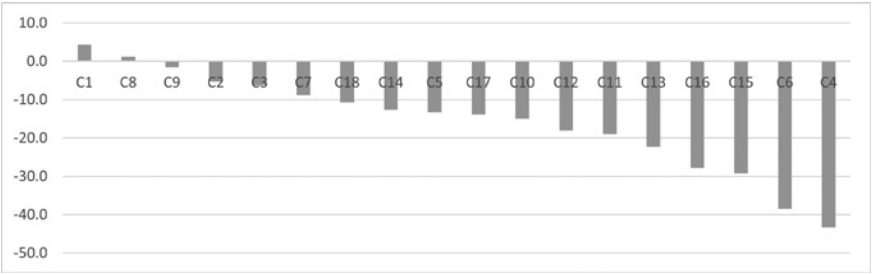


FIGURE 8 Change in item facility for computation tasks

Figures 7 and 8 depict the differences in item facilities between the English students' anno 1976 and the Danish students' anno 2023. A negative score indicates a higher item facility in England 1976.

It is evident that the item facilities for 12 out of the 18 computation items have decreased by over 10%. In comparison, this is only the case with 7 out of the 41 word-items, as depicted in Figures 7 and 8.

Further, the item facilities for nine items increased by over 5% for the Danish students compared to the English students (see Table 9). The items are presented in decreasing order due to the change in item facilities.

Generally, the increase is seen in tasks that involve shading of figures and labelling of fractions when figures are involved. However, the decrease in item facilities (Table 10) is much more significant.

The decreasing items involve the computation of fractions involving mixed numbers, thereby indicating that the Danish students are not used to such computations. The biggest drop in item facility within the word problems is seen in task P18 (Table 10). In both P18a and P18b, the proportion of different error types have decreased; thus, the decrease in item facility is solely due to an increase in missing responses from 7% to 33% and from 10% to 36%, respectively. This could indicate that the question was rather difficult for the Danish students to understand. This was either due to a poor translation or because

TABLE 9 Item facilities for items with increasing facilities

| Item | Item description | England 1976 | Denmark 2023 | Change |
|---|---|---|---|---|
| P11 | Choose the box with the most shaded.  | 67% | 78% | 11% |
| P8a | What fraction is shaded? (Showing four of eight in a rectangle) | 79% | 86% | 7% |
| P6a | What fraction is cracked when 5 of 12 eggs are cracked? | 73% | 79% | 7% |
| P7 | Tick each shape in which half has been shaded.  | 76% | 82% | 6% |
| P4c | Shade two-thirds of six. | 64% | 70% | 6% |
| P4d | Shade two-thirds of nine. | 64% | 70% | 6% |
| P22 | What fraction of the whole disc is shaded when one-sixth of three-fourths is shaded? | 52% | 57% | 5% |
| P21a | What fraction does John have left if he pays $\frac{3}{5}$ in tax? | 72% | 77% | 5% |
| P16a | $\frac{2}{7}=\frac{?}{14}$ | 55% | 60% | 5% |

the context was unfamiliar to the students. Further, there are six items with small or no changes in facilities, which are categorised into 1) tasks with high facilities in both decades and 2) tasks with low facilities in both decades.

### 4.5 *Equivalent Fractions*

When examining items where students are to find a missing value to make two fractions equivalent, changes are mainly seen in the first three items (see Table 11). In the first two items, there is a decrease in the item facilities at 11%. However, the facility for the third item increases with 12%.

Even though the first two items involve doubling or halving, the Danish students perform better on items that involve a multiple of 10. This type of task is not identified in Danish textbooks from either decade; however, even though these tasks may not be familiar to the Danish students, between 75%

TABLE 10 Item facilities for items with decreasing facilities

| Item | Item description | England 1976 | Denmark 2023 | Difference |
|---|---|---|---|---|
| C4 | $1-\frac{5}{12}$ | 65% | 22% | 43% |
| C6 | $3\times 10\frac{1}{2}$ | 82% | 43% | 38% |
| P18a | How many spokes $10\frac{1}{2}$ long can be cut from a 40 cm long wire? | 74% | 44% | 30% |
| C15 | $1\frac{1}{5}-\frac{3}{5}$. | 52% | 22% | 29% |
| P5 | What is 17 cm divided into four equal pieces? ($4\frac{1}{4}$) | 54% | 26% | 29% |
| C16 | $2\frac{1}{2}-\frac{3}{4}$ | 42% | 14% | 28% |
| C13 | $2\times\frac{1}{8}$ | 48% | 26% | 22% |
| P18b | What length of wire is left after cutting three spokes of $10\frac{1}{2}$ cm? | 56% | 35% | 21% |

TABLE 11 Equivalence (Items P14a–e, P16a–b)

| | $\frac{1}{3}=\frac{2}{?}$ | $\frac{6}{8}=\frac{3}{?}$ | $\frac{5}{10}=\frac{?}{30}$ | $\frac{2}{3}=\frac{?}{15}$ | $\frac{4}{12}=\frac{1}{?}$ | $\frac{2}{7}=\frac{?}{14}$ | $\frac{2}{7}=\frac{n}{14}=\frac{10}{?}$ |
|---|---|---|---|---|---|---|---|
| Danish students | 66% | 58% | 70% | 60% | 52% | 60% | 26% |
| English students | 77% | 69% | 58% | 58% | 52% | 55% | 24% |

and 80% of the students attempted to give an answer, except in the seventh item (Table 11, P16b), with 34% missing answers. The reason for this change may be explained by the different strategies students use to answer unknown tasks indicated by the error codes presented in Table 12.

It appears that students attempt to find a pattern of some kind when encountered with such an unfamiliar task. However, the type of pattern varies depending on the numbers given. In all P14 items, one of the strategies used by students' is to copy either the denominator or the numerator from the known fractions. Generally, students utilise different additive strategies to find a pattern, but this is not the case in the third item, where there are very few wrong answers. It appears that the students who use different additive strategies in

TABLE 12 Student errors in unknown equivalence tasks (percentage)

| | $\frac{1}{3}=\frac{2}{?}$ | $\frac{6}{8}=\frac{3}{?}$ | $\frac{5}{10}=\frac{?}{30}$ | $\frac{2}{3}=\frac{?}{15}$ | $\frac{4}{12}=\frac{1}{?}$ | $\frac{2}{7}=\frac{?}{14}$ | $\frac{2}{7}=\frac{n}{14}=\frac{10}{?}$ |
|---|---|---|---|---|---|---|---|
| $\frac{a}{b}=\frac{a\pm n}{b\pm n}$. | 2% | 11% | | | 2% | 2% | |
| $\frac{a}{a\pm n}=\frac{b}{b\pm n}$. | | | 2% | 3% | | | 4% |
| $\frac{a}{b}=\frac{c}{b\pm c}$. | 5% | | | | | | |
| Copy either numerator or denominator from the known fraction | 3% | 1% | 1% | 2% | 2% | | |
| Other errors | 4% | 7% | 2% | 12% | 20% | 13% | 36% |

the other similar items select a multiplicative strategy involving the table of 10, thereby yielding a correct response. In the last item, students tend to overgeneralise a pattern identified when going from the first to the second fraction, using the table of two for the numerator and the table of seven for the denominator. Further, 7% of the other errors are students who answer with 21 and 28; 42 was also part of student errors, but this response does not have a separate code.

### 4.6 *Mathematical Argumentation*

Item P17 (Table 13) is one out of three items where code 1 is only given if the student both answers yes/no and provide a justification for their answer; the other two are items P3 and P13.

In both England and Denmark, the item facility is very low, at only 2%, even though most students attempt to answer the question. The answer 'Yes, because $\frac{1}{4}$ is greater than $\frac{1}{2}$', did not exist in the original marking key; thus, if English students answered in this manner, these answers would be part of code 9 in England. The two most typical answers given by both Danish and English students is divided between students answering 'Yes, because Mary has more' and 'No, because $\frac{1}{2}$ is greater than $\frac{1}{4}$'. While the latter indicates that these students only compare the fractional number without considering the whole the fraction is to be seen in relation to, the first type is typical for students who know that the fractional number is a fraction of something. Here, typical student answers are different variations of 'Yes, because we do not know how much money they have each' and 'Yes, because Mary could have more money'. This is

TABLE 13 P17. Mary spends $\frac{1}{4}$ of her pocket money, and John spends $\frac{1}{2}$ of his. Is it possible for Mary to have spent more than John? Why do you think this?

| | **Yes, if Mary has (at least) twice as much** | **Yes (gives a correct numerical example)** | **Yes, because $\frac{1}{4}$ is greater than $\frac{1}{2}$.** | **Yes, because Mary has more** | **Yes (no reason)** | **No, because $\frac{1}{2}$ is greater than $\frac{1}{4}$ (or no reason)** |
|---|---|---|---|---|---|---|
| England | 2% | 4% | -– | 39% | 12% | 34% |
| Denmark | 2% | 5% | 2% | 34% | 5% | 41% |

TABLE 14 P13. Peter and Abdul each have a bar of chocolate of the same size. Peter breaks his into eight equal pieces and eats four. Abdul breaks his into four equal pieces and eats two

| | **The same because $\frac{4}{8}=\frac{2}{4}$ (or no reason)** | **Abdul because $\frac{2}{4}>\frac{4}{8}$** | **Peter because $\frac{4}{8}>\frac{2}{4}$** |
|---|---|---|---|
| England | 64% (–) | 10% | 23% |
| Denmark | 59% (12%) | 11% | 12% |

also the case for students who provide a correct numerical example like 'Yes, if she had 10 000 and he only had 10'. Very few of the Danish students provide an answer to this question without giving some kind of justification. However, not many of the arguments given are mathematical arguments, which indicates that even though Danish students are used to arguing/justifying in the mathematical classroom, coming up with a mathematical justification appears to be rather difficult for them. Another interpretation could be that the question is not formulated in a manner that guides students to come up with a mathematical answer. The latter interpretation is supported by the much higher item facilities for P13 (Table 14).

In the English results, it appears as though all answers identifying that the two boys eat the same would be coded as correct. In the Danish coding manual, a code was added to capture the 12% who could not formulate a mathematical explanation — for example, 'They eat the same because they eat the same' or 'They eat the same because I know so'. Few of the Danish students wrote $\frac{4}{8}=\frac{2}{4}$ directly. Instead, they argued saying, 'Because 4 is half of 8 and 2 is half of 4' or 'Because they both eat half of what it was divided into'. Even though the item

facilities in both countries indicate that the students are capable of making a mathematical argument using equivalent fractions, the fact that so many of the Danish students argue by reference to one half indicates that it could be interesting to see if the students would be able to argue using equivalence if the fractions did not correspond to one half.

### 4.7 *Fractions and Decimal Numbers*

In total, there are six questions where the Danish students are prone to answer using decimals, as shown in Table 15. Common to them all is that the questions include computation with whole numbers or mixed numbers involving half and fourths.

C1 is the only computation task with a clear increase in item facility compared to the original data, increasing from 31% in 1976 to 35% in 2023. When closely examining the Danish answers, 90% of the correct answers by the Danish students are expressed in decimals. We do not know how many of the English students answered using decimals in 1976, but the increase in item facility despite a 26% increase in missing responses could indicate that the Danish students are more likely to answer using decimal numbers. Further, the fact than only 12 of the 336 Danish students answered this question by the fraction $\frac{3}{5}$ could indicate that Danish students generally do not think of a fraction as a legitimate answer in this context. Moreover, there is no indication in the Danish fraction instruction that the students would ever encounter such a type of task in the Danish textbooks. The corresponding problem item, 'Three plates of chocolate is to be divided equally between five children. How much should each child get?', even though answered correctly by fewer Danish students, reveals that 51% of the correct answers are in decimals, 5% in percentage, and 44% in the fraction $\frac{3}{5}$. This indicates that the context in which the question is set determines whether the students perceive a fraction as a legitimate answer, an error that also appears to depend on the context in which the given question involves remainder solutions. Where the English students in 1976 used remainder solutions to solve computation tasks — for example., '1 remainder 2' in task C1 or '3 remainder 3' in C3, identified in 9.4% and 17.2% student answers, respectively — this is barely seen (e.g., 1% and 0%, respectively) in the Danish results. However, 15% of the Danish students use a remainder solution in P15 and cover different versions of 'One half each and one half left for tomorrow/my mother'. This type of answer was not coded in the English result; however, they could have been part of the 40% code 9 responses. Since the marking key has three other error codes involving remainder solutions, it does however not appear likely that they would not involve remainder solutions in the coding of item P15. The general lack of remainder

TABLE 15 Use of decimal answer

| Item | Item description | Correct (fraction answer) | Correct (decimal answer) |
|---|---|---|---|
| P15 | Divide three bars of chocolate equally between five children. $\frac{3}{5}$ | 10% | 11% |
| P18b | What length of wire is left after cutting three spokes of $10\frac{1}{2}$ cm? | 24% | 11% |
| C1 | $3 \div 5$ | 4% | 32% |
| C3 | $15 \div 4$ | 4% | 26% |
| C6 | $3 \times 10\frac{1}{2}$ | 17% | 26% |
| C11 | $1\frac{1}{4} \times 7$ | 16% | 9% |

solutions in the Danish results are probably due to the changes in curriculum since division focusing on remainders does not get much focus in Danish textbooks nowadays.

In C6, students are asked to do a multiplication of a whole number and a mixed number: $3 \times 10\frac{1}{2}$. In England, almost all students attempted to answer this question and there is an item facility of a little over 80%. The 39% decrease in item facility is partially explained by an increase of 29% in missing responses. This indicates that this type of task is less familiar to Danish students in 2023 than it was for English students in 1976. The remaining difference is divided between codes 8 and 9. Code 8 responses are $30\frac{1}{2}$, thereby indicating that students multiply each integer by 3. Approximately 30% of these student's respond with $30\frac{3}{6}$ and 15% with the decimal equivalent. This type of error has almost doubled, leaving approximately 10% of the Danish students. Among the 13% code 9 responses are 30 and 45. The answer 30 can be interpreted as a student disregarding the fraction in the mixed number and only multiplying the two whole numbers. The 45 is interpreted as $3 \times 15$, which is the result of students reading $10\frac{1}{2}$ as '10 and half of 10'. Both types of responses indicate that at least some of the Danish students have trouble giving meaning to mixed numbers, which was a lot less common among English students in 1976. Moreover, 60% of the correct Danish answers are given using decimal numbers, thereby converting to $3 \times 10.5 = 31.5$.

### 4.8 *Common Errors*

In the CSMS tests, common errors were coded in order to identify typical errors at the school or student level. For 16 of the items, specific errors accounted for

over 10% of all responses and, consequently, were judged to indicate potential difficulties or possible misconceptions faced by students; this is in comparison to 21 items in England in 1976. The decrease in the use of error codes, compared with a decrease in item facilities for many of the items, indicates that the Danish students are more prone to skip an item if they find the item difficult, which makes a diagnostic test less informative. In the Problem part of the test, these errors are typically seen in multiple choice items, equivalence (accounted for in equivalent fractions), tasks involving shading of parts (accounted for in overall performance), and tasks involving argumentation (see mathematical argumentation). An example of a multiple-choice item is P5, where the item facilities decreased from 54% to 26%.

> A piece of ribbon 17 cm long has to be cut into 4 equal pieces.
> Tick the answer you think is most accurate for the length of each piece.
>
> (a) 4 cm, remainder 1 piece
> (b) 4 cm, remainder 1 cm
> (c) $4\frac{1}{4}$ cm
> (d) $\frac{4}{17}$ cm

The number of students who believed that the answer should be an equal number increased from 30% to 38%. However, there is a decrease in students who answered (a), thereby indicating that Danish students are more likely to answer with a whole number and a correctly identified entity than 'one piece'. The largest proportional change, not counting in the missing answers, is seen in (d), which increased from 7% to 16%, thereby indicating that a larger proportion of the Danish students have difficulties identifying the correct fractions, believing that the numerator should always be smaller than the denominator.

In both England and Denmark, 20–25% of the students answer using a fraction in the numerator in item P20, as depicted in Table 16. We do not know if the picture would have been the same in Denmark in the 1970s, where at least some of the textbooks had examples of what did not constitute a fraction. However, despite the time difference, there are almost as many students who would write the numerator using a fraction as there are students who would use whole numbers in both the numerator and denominator. This could indicate that the manner in which the students write the answer depends on what they identify as the whole. Students that count using the shaded triangles would probably use the first example as their answer, whereas the second answer are given by students that uses the white squares as the whole. Consequently, we do not know what would constitute a fraction for the first group. This type of

TABLE 16 P20. What fraction of the floor has been tiled? Tiles are shaded 

| | $\frac{9}{24}$ | $\frac{4\frac{1}{2}}{12}$ |
|---|---|---|
| England | 29% | 22% |
| Denmark | 28% | 23% |

TABLE 17 Overgeneralisation of whole number knowledge: $\frac{a}{b}+\frac{c}{d}=\frac{a+c}{b+d}$

| | **England** | **Denmark** |
|---|---|---|
| C8. $\frac{3}{8}+\frac{2}{8}$ | 20% | 7% |
| C14. $\frac{1}{3}+\frac{1}{4}$ | 29% | 17% |
| C18. $\frac{2}{7}+\frac{3}{4}$ | 30% | 16% |

TABLE 18 Overgeneralisation of whole number knowledge: $a\cdot\frac{b}{c}=\frac{a\cdot b}{a\cdot c}$

| | **England** | **Denmark** |
|---|---|---|
| C13. $2\cdot\frac{1}{8}$ | 17% | 23% |

error is seen in different versions in three additional items among the Danish students; however, this is not as widespread as in P20.

In the Computation test, typical errors had to do with overgeneralisation of whole number knowledge (Ni & Zhou, 2005), where students see the numerator and denominator as two separate numbers, adding numerator and denominator as if they were whole numbers.

For the Danish students, this type of error mainly occurs when the denominators are different as seen in Table 17. When multiplying a whole number with a fraction there is an increase in students multiplying the whole number with both the numerator and the denominator (Table 18).

This type of error can either be a consequence of repeated instruction in finding equivalent fractions where students do not realise that equivalent fractions are simple different representations of the same rational number. Alternatively, this can again be seen as an overgeneralisation of whole number

knowledge, thereby leading the students to multiply both denominator and numerator by the whole number.

## 5 Discussion

In this paper, we reported on a conceptual replication of Fractions 1. The test was originally used in the 1970s in England to investigate students' understanding of fractions and subsequently as a diagnostic test for use by teachers (Hart et al., 1985). We conducted a scaling-out replication of the original study with a different population and at a different time (Danish students in 2023 rather than the original English student population in 1976, nearly half a century earlier). The aims of our replication were to assess the extent to which the original findings still hold true and to describe any changes. Specifically, we sought to (re-)validate the test in this new and current context. We investigated the value of the test as a diagnostic tool for Danish teachers. Finally, we aimed to provide evidence of the strengths and weaknesses of Danish 12–14-year-old students in comparison to English 12–13-year-old students in the 1970s and, thus, assess the extent to which the original study's widely cited findings about students understanding of fractions remain valid.

We first revalidated the test to assess its value as an instrument for investigating students' understanding of fractions in today's classrooms in Denmark, and, thus, for a comparison with the original English sample of 12–13-year olds some 40 years previously.

### 5.1 *The CSMS Fractions 1 as a Valid Assessment of Students' Knowledge of Fractions*

In the current study, the validity of the Fractions 1 test is determined by including an extension (Brown & Wood, 2018) to the original study using Rasch analysis. The Rasch analysis indicated a good fit to the model, with infit measures within the productive range. The nine outfitting items have values that are not considered a threat to the validity of the test. Overall, the test performs at least as well as that in the original study. Furthermore, according to the Rasch model, the Wright Map identifies four clusters of tasks ranging from easy items of counting and labelling tasks, which can be solved using only whole number knowledge. The second cluster involves labelling of fractions using discrete units, equivalence, items focusing of the size of a fraction, and addition of fractions. The third cluster of tasks involved items with more than one operation, where finding equivalent fractions was often necessary to compute an answer.

The fourth cluster involves computational items with either division of fractions or subtraction involving mixed numbers.

### 5.2 *The CSMS Fractions 1 as a Tool in Teachers' Formative Assessment*

Since the CSMS tests were designed as a diagnostic tool for teachers in formative assessments, the four levels of understanding were developed to communicate to teachers a broad sense of how different students understand fractions. As indicated in both the clustering of items in the Rasch analysis and the item facilities for the Danish students, and when examining the mathematical ideas students are expected to engage with when solving the tasks, the CSMS levels 2 and 3 overlap in the current data. This is opposed to the original data, where the item facilities were clearly separable. These changes in item facilities are identified partially due to a decrease in the performance of Danish students on items related to computation of equivalent fractions by doubling and partially due to an increase in Danish performance related to labelling of fractions, both using continuous and discrete data, and colouring of fractional parts. Even though most items in the CSMS Fractions 1 test correspond to the fraction instruction seen in most (Danish) schools (Haahr & Jensen, 2008), or at least the Danish textbooks, there are tasks that are no longer relevant. This applies both to tasks that the Rasch analysis identifies as not contributing to knowledge about the students' understanding of fractions as well as to tasks that are no longer part of the curriculum — for example, computation using mixed numbers — and tasks where the context is unfamiliar to students (e.g. using bicycle spokes or marbles). Furthermore, the test includes tasks where a wrong strategy can lead to the correct answer. Therefore, the items could benefit from being changed, thereby leading to more precise identification of students' errors.

For the majority of the level 2 and level 3 items, it becomes difficult to separate the two levels not least because of the many different fraction ideas involved at both levels. These items involve both colouring of figures using part of a whole interpretation where students need to realise that two parts should be shaded for every three parts in the figure, the order of unit fractions, addition of fractions with common denominator, and equivalence items. With so many different ideas present at one level, it can be discussed what knowledge teachers actually gain about students with test scores corresponding to levels 2 and level 3. However, if clustered in minor, clearly separable pieces around the mathematical ideas involved, they could prove useful in teachers' formative assessment of students' knowledge of fractions, regardless of the curriculum at a specific time in a specific country. Thus, it makes sense to acknowledge

that fraction instruction might not be hierarchical in structure but consists of several interrelated strand of key ideas (Confrey et al., 2009).

### 5.3 *Strengths and Weaknesses of Danish Students' Performance Compared to that of English Students in the 1970s*

The present conceptual replication (Schmidt, 2009) identifies numerous similar findings as the CSMS — this is despite scaling out (Melhuish, 2018) by examining a different population almost 50 years later. As in CSMS (Hart et al., 1989), the Danish students perform well on items involving labelling of fractions, whereas addition and multiplication of fractions proves much more difficult. Danish students appear to be slightly stronger in labelling of fractions and colouring fractional parts. These tasks depend on an interpretation of a fraction as a part of a whole, which is strongly emphasised in Danish textbooks (Færch & Pedersen, 2023) and they can be solved solely using knowledge of whole numbers. In both countries, students avoid using fractions (Hart, 1981) in computation tasks. However, while English students do so by finding a solution involving remainders, the Danish students convert the solution to decimals. Both the Rasch analysis and the differences in item facilities indicate that the Computations part of the test posed greater challenges, particularly for the Danish students than the first part of the Fractions test involving problems using words or diagrams. Generally, it appeared as though the more complex the fraction item was, the more students applied misremembered rules for fraction computation. The analysis does not reveal any major differences in items where students are to find equivalent fraction. Although most equivalence tasks are unfamiliar to the Danish students (see Section 2.2: Development of Fraction Instruction in Denmark Since the 1970s), it appears as though Danish students look for patterns when they do not know how to solve a task. The students search for a pattern causes them to behave unexpectedly which, for example, leads to higher facilities in equivalence tasks involving a multiple of 10 compared to doubling. This results in applying additive strategies when the equivalence task involves smaller numbers and multiplicative strategies when the numbers are larger, resulting in a correct answer even though the doubling task would be expected to be less difficult than the others. However, when items involve the computations of fractions where finding a common denominator (i.e., finding an equivalent fraction is necessary), there is a major decrease in the Danish facilities indicating that numerous students do not know the meaning of equivalence — for example, they apply a rule they remember when asked directly for it but cannot apply the knowledge in more complex situations. This was also the case for the English

students (Kerslake, 1986). Further, Danish students appear to have difficulties with mixed numbers, either due to difficulties in giving meaning to these — for example, believing that $10\frac{1}{2}$ is equal to 15 (10 and half of 10) or due to whole number bias (Nunes & Bryant, 1996) — thereby leading students to operate on the whole number, the numerator, and the denominator as three separate numbers. In addition, many Danish students convert from mixed numbers to decimal numbers before performing the computation. As in CSMS (Hart, 1981), the Danish students have trouble understanding a fraction as a rational number, thereby limiting the Danish students' possibilities to use fractions in more complex computations and understand the meaning of mixed numbers.

The distribution of the missing items in the computations tasks indicates that the students did not simply stop due to time limitations or other external factors. Even though more students generally attempted to answer more items on the first computation page than on the last, the fact that C8 has the lowest proportion of missing items could be an indication than the Danish students looked through the items and only answered the ones they found to be the easiest or most familiar.

### 5.4 *The Validity of the CSMS Findings on Students' Understanding of Fractions*

The results from the current replication study indicates, that although the CSMS Fractions 1 test continues to be a valid measurement of students' understanding of fractions, there are elements in the design of the test and the corresponding levels of understanding that means that, to function as a diagnostic tool, the test needs some adaptation in order to be relevant to other countries or new curricula. In part, the CSMS findings on fractions must be seen in relation to the English students and the existing curriculum with a focus on fractions and mixed numbers. It could be, for example, the fact that English students in the 1970s used remainders to answer more challenging tasks, while students in Denmark in the 2020s more often convert to decimal numbers — even if it is not always meaningful. In part, the CSMS levels of understanding were developed in the context of the mathematics curriculum in England in the 1970s. Since the relative difficulties of items are undeniably linked to the student's previous mathematical experiences, we see in the present study that Danish students, for example, look for patterns when they encounter a type of task unknown to them. Since the Danish students are used to looking for number patterns, but do not seem to have developed an understanding of fractions as rational numbers, the item difficulty for some of the equivalence items is changed — both compared to the order of the English students and what would be expected from a theoretical perspective.

## 6 Conclusion

In the current study, we replicated the CSMS Fractions 1 test originally developed in England in the 1970s. This is a seminal study that is widely cited as providing evidence of how students around the world understand fractions. However, the original study was conducted nearly 50 years ago with a sample of students from a particular educational system, England, and was not able to make use of the more sophisticated statistical methods that have been developed over the intervening half century. Moreover, since 1976 there has been a substantial body of research on the teaching and learning of fractions (e.g., see Lamon, 2007, 2012) and it is possible that changes to teaching methods and curriculums since the 1970s may have had an impact on how students understand fractions. Hence, our replication study sought to re-validate the test, to assess its value as a diagnostic tool for teachers and to examine whether the original results still hold by comparing the strengths and weakness of the current Danish students to those of the original sample of English students five decades ago.

To do this, we conducted a Rasch analysis to re-validate the CSMS Fractions 1 test. In general, this analysis indicates that the test provides a meaningful measure of student abilities and, in contrast to the original study, we found little evidence to suggest that any items should be excluded from the scale. Indeed, our analysis indicates that the computational items, which were originally excluded, could all be included in the main scale. However, our analysis indicates some important changes to the original hierarchical structure of the levels of understanding, where, for current Danish students, the original level 2 and level 3 appear to cluster together. As previously noted, at level two, the students understanding of fractions was limited to discrete quantities, obtaining equivalent fractions by doubling, and perform addition of two fractions with the same denominator. In contrast, at level three, students had a broader understanding of equivalence and ordering. The clustering indicates a more iterative view of the development of equivalence than was evidenced in English sample in the 1970s.

Despite this change, the hierarchy of levels appeared to continue to be a useful approach to summarising the progression in student understanding. A comparison across both countries in the different time periods reveals a roughly equivalent distribution of students across the levels of understanding. However, this obscures some major lower item facilities for the Danish students, particularly in items involving computations with fractions in situations where part of a whole interpretation and whole number counting strategies are not sufficient. The Danish students do appear to be more familiar with

labelling of fractions, regardless of continuous or discrete entities, thereby resulting in increases in item facilities compared to the original study. Despite these positive findings, our result indicates that the majority of Danish students have trouble understanding fractions as rational numbers, have trouble with fraction computation, and cannot reason with fractions. This was also a problem in the 1970s, but our results reveal that it is now the case for a larger proportion of students. In Denmark, there is far less focus on fractions and rational numbers today than there was in the textbooks of the 1970s. When the current fraction instruction leads to few students being able to reason with fractions or understand a fraction as a rational number, we need to take a critical look at the current focus in fraction instruction and look for alternatives.

In summary, our replication study provides a valuable contribution in updating the findings of the CSMS study in a new and current context. We have shown that many of the findings of the original study broadly hold true, but there are some significant differences. In particular, we have shown that a test developed almost 50 years ago in England, is still a valid and useful tool for assessing students understanding of fractions, although the test would need adaptation for use as a diagnostic tool for teachers. As a final comment, we emphasise the importance of conducting replication studies of the kind that we have carried out. Many of the seminal studies in mathematics education were conducted many decades ago in specific contexts and with less sophisticated methods than are available today. By replicating these studies, research can re-validate the methods and re-assess the original findings, but research can also, as we have done in this study, add nuance to the original findings. In doing so, replication has a vital role in potentially strengthening, or challenging, key findings in mathematics education.

**Acknowledgements**

We thank Dr. Dietmar Küchemann for fruitful discussions during the preparation of this paper.

**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.

Aksu, M. (1997). Student performance in dealing with fractions. *The Journal of Educational Research*, *90*(6), 375–380. https://doi.org/10.1080/00220671.1997.10544595.

Andersen, M. W., Lindhardt, B., & Dalsgaard, R. S. (2016). *Kontext+ 6*. Alinea.

Andersen, M. W., Lindhardt, B., Dalsgaard, R. S., & Hessing, S. (2015). *Kontext+ 5*. Alinea.

Andersen, M. W., Lindhardt, B., Dalsgaard, R. S., & Poulsen, M. (2014). *Kontext+ 4*. Alinea.

Baturo, A. R. (2004). Empowering Andrea to help year 5 students construct fraction understanding. In M. J. Høines & A. B. Fuglestad (Eds.), *Proceedings of the 28th PME Conference* (Vol. 2, pp. 95–102). Bergen University College.

Behr, M. J., Lesh, R., Post, T. R., and Silver, E. A. (1983). Rational number concepts. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and processes* (pp. 91–125). Academic Press.

Behr, M. J., Wachsmuth, I., Post, T. R., & Lesh, R. (1984). Order and equivalence of rational numbers: A clinical teaching experiment. *Journal for Research in Mathematics Education*, *15*(5), 323–341. https://doi.org/10.5951/jresematheduc.15.5.0323.

Bisgaard, A. S., Bull, A. R., Hansen, H. H., Kreiberg, C., & Skipper-Jørgensen, A. (2015). *Multi 7*. Gyldendal.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental measurement in the human sciences* (2nd ed.). Lawrence Erlbaum.

Brown, A. N. & Wood, B. D. K. (2018). Which tests not witch hunts: A diagnostic approach for conducting replication research. *Economics: The Open-Access, Open-Assessment E-Journal*, *12*(53), 1–26. https://doi.org/10.5018/economics-ejournal.ja.2018-53.

Brown, M. (1996). The context of the research: the evolution of the national curriculum for mathematics. In D. C. Johnson & A. Millett (Eds.), *Implementing the mathematics national curriculum: Policy, politics, and practice* (pp. 1–28). Paul Chapman Publishing.

Brown, M., Bibby, T., & Johnson, D. C. (2000). Turning our attention from the what to the how: The national numeracy strategy. *British Educational Research Journal*, *26*(4), 457–471. https://doi.org/https://doi.org/10.1080/713651570.

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). BSRLM.

Carraher, D. W. (2013). Learning about fractions. In L. P. Steffe & P. Nesher (Eds.), *Theories of mathematical learning* (pp. 253–278). Lawrence Erlbaum.

Cascella, C., Giberti, C., & Maffia, A. (2023). Percentages versus Rasch estimates: alternative methodological strategies for replication studies in mathematics education. *Research in Mathematics Education*, 1–19. https://doi.org/10.1080/14794802.2022.2154826.

Clausen, C. C. K., & Jensen, C. E. (1971). *Matematiktimen 4. Grundbog*. Gjellerup & Gad.

Clausen, C. C. K., & Jensen, C. E. (1975). *Matematiktimen 5. Grundbog*. Gjellerup & Gad.

Clausen, C. C. K., & Jensen, C. E. (1976). *Matematiktimen 6. Grundbog*. Gjellerup & Gad.

Clausen, C. C. K., & Jensen, C. E. (1977). *Matematiktimen 7. Grundbog*. Gjellerup & Gad.

Confrey, J., Maloney, A. P., Nguyen, K. H., Mojica, G., & Myers, M. (2009). Equipartitioning/splitting as a foundation of rational number reasoning using learning trajectories. In M. Tzekaki, M. Kaldrimidou, & H. Sakonidis (Eds.), *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 345–352). PME.

Cort, J., & Johannessen, E. (1970). Matematik 5. Grundbog. Gjellerup & Gad.

Cort, J., & Johannessen, E. (1973). Matematik 6. Grundbog. Gjellerup & Gad.

Cort, J., & Johannessen, E. (1974). Matematik 4. Grundbog. Gjellerup & Gad.

Cort, J., & Johannessen, E. (1975). Matematik 7. Grundbog. Gjellerup & Gad.

Technical University of Denmark (DTU) (2021). *Transportvaneundersøgelsen. Faktaark om børn og unges transportvaner* (The transport habits survey. Fact sheet on children and young people's transport habits). https://www.cta.man.dtu.dk/-/media/centre/modelcenter/2021_faktaark-boern-og-unge.pdf.

Færch, J. V., & Pedersen, P. L. (2023, July 10–14). *Representation of fractions in Danish fourth-grade textbooks* (Paper presentation). 13th Congress of the European Society for Research in Mathematics Education (CERME13), Budapest, Hungary.

Gabriel, F., Coché, F., Szucs, D., Carette, V., Rey, B., & Content, A. (2013). A componential view of children's difficulties in learning fractions. *Frontiers in Psychology*, *4*, Article 715. https://doi.org/10.3389/fpsyg.2013.00715.

Haahr, O. & Jensen, H. N. (2008). Folkeskolens og læreruddannelsens regne- og matematikundervisning (Arithmetic and mathematics teaching in public schools and teacher training). In *Matematikundervisningen i Danmark i 1900-tallet. Grundlæggende regning og matematik*. Bind 1 (pp. 163–406). Syddansk Universitetsforlag.

Hambleton, R. K. (1993). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Educational measurement* (Vol. 3rd ed., pp. 147–200). Oryx Press.

Hannula, M. S. (2003). Locating fraction on a number line. In N. A. Pateman, B. Dougherty, & J. Zilliox (Eds.), *27th International Group for the Psychology of Mathematics Education Conference* (pp. 17–24). PME. https://files.eric.ed.gov/fulltext/ED500981.pdf.

Hansen, N. J., Johnsen, L., Lindhardt, B., & Thomsen, H. (2015). *Kontext+ 7*. Alinea.

Hart, K. (1981). Fractions. *Mathematics in School*, *10*(2), 13–15.

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

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

Hart, K. M., & Johnson, D. C. (1984). *Secondary school children's understanding of mathematics. A report of the mathematics component of the concepts in secondary mathematics and science programme*. Chelsea College; University of London; Centre for Science Education.

Hart, K. M., Johnson, D. C., Brown, M., Dickson, L., & Clarkson, R. (1989). *Children's mathematical frameworks 8–13: A study of classroom teaching*. NFER-Nelson.

Helsted, H. (1975). *Lov om folkeskolen af 26. juni 1975 med indledning og kommentarer* (Act on the primary school of 26 June 1975 with introduction and comments). Finn Suenson Forlag.

Hodgen, J., Foster, C., & Brown, M. (2022). Low attainment in mathematics: An analysis of 60 years of policy discourse in England. *The Curriculum Journal, 33*(1), 5–24. https://doi.org/https://doi.org/10.1002/curj.128.

Hodgen, J. (2012). Computers good, calculators bad. In P. Adey & J. Dilllon (Eds.), *Bad education: Debunking educational myths*., Open University Press.

Jensen, H. N., Lindhardt, B., & Møller, M. T. (2016). *Kontext+ 3b*. Alinea.

Kerslake, D. (1986). *Fractions: children's strategies and errors. A report of the strategies and errors in secondary mathematics project*. NFER-Nelson.

Kieren, T. E. (1976). On the mathematical, cognitive, and instructional foundations of rational numbers. In R. Lesh (Ed.), *Number and measurement: Papers from a research workshop* (pp. 101–144). The Georgia Center for the Study of Learning and Teaching Mathematics; Department of Mathematics Education University of Georgia.

Kieren, T. E. (1988). Personal knowledge of rational numbers: Its intuitive and formal development. In J. Hiebert & M. Behr (Eds.), *Research agenda for mathematics education: Number concepts and operations in the middle grades* (Vol. 2, pp. 162–181). Lawrence Erlbaum.

Kieren, T. E. (1993). Rational and fractional numbers: From quotient fields to recursive understanding. In T. P. Carpenter, E. Fennema, and T. A. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 49–84). Erlbaum.

Kristensen, B., & Teglskov, R. (2012). *Multi 3b*. Gyldendal.

Lathi, S. S., Mogensen, P., Sperling, P. H., & Westfall, B. (2016). *Multi 5*. Gyldendal.

Lamon, S. J. (1999). *Teaching fractions and ratios for understanding*. Lawrence Erlbaum.

Lamon, S. J. (2007). Rational numbers and proportional reasoning: Towards a theoretical framework for research. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 629–667). NCTM.

Lamon, S. J. (2012). *Teaching fractions and ratios for understanding. Essential content knowledge and instructional strategies for researchers* (3rd ed.). Routledge.

Lin, F.-L. (1988). Societal differences and their influences on children's mathematics understanding. *Educational Studies in Mathematics, 19*(4), 471–497. https://doi.org/10.1007/BF00578695.

Linacre, J. M. (2002). What do infit and outfit, mean-square, and standardized mean? *Rasch Measurement Transactions*, *16*(2), 878. https://www.rasch.org/rmt/rmt162f.htm.

Mair, P., Hatzinger, R., & Maier, M. J. (2021). *eRm: Extended Rasch modelling* (Version 1.0–2) (Computer software). https://cran.r-project.org/package=eRm.

Melhuish, K. (2018). Three conceptual replication studies in group theory. *Journal for Research in Mathematics Education*, *49*(1), 9–38. https://doi.org/10.5951/jresematheduc.49.1.0009.

Millett, A., & Johnson, D. C. (2000). Dancing to whose tune? Conflicting priorities for primary schools. *Education 3–13*, *28*(3), 24–31. https://doi.org/10.1080/03004270085200291.

Mogensen, P., Teglskov, R., & Westfall, B. (2014). *Multi 6*. Gyldendal.

Mogensen, P., Westfall, B., & Sperling, P. H. (2015). *Multi 4*. Gyldendal.

Ni, Y., & Zhou, Y. D. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist*, *40*(1), 27–52. https://doi.org/10.1207/s15326985ep4001_3.

Nunes, T., & Bryant, P. (1996). *Children doing mathematics*. Blackwell.

Pedersen, P. L., Aunio, P., Sunde, P. B., Bjerre, M., & Waagepetersen, R. (2022). Differences in high-and low-performing students' fraction learning in the fourth grade. *The Journal of Experimental Education*, *91*(4), 636–654. https://doi.org/10.1080/00220973.2022.2107603.

Pitkethly, A., & Hunting, R. (1996). A review of recent research in the area of initial fraction concepts. *Educational Studies in Mathematics*, *30*(1), 5–38. https://doi.org/10.1007/BF00163751.

Putra, Z. H. (2018). *A comparative study of Danish and Indonesian pre-service elementary teachers' knowledge of rational numbers*. Department of Science Education, Faculty of Science, University of Copenhagen.

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90–100. https://doi.org/10.1037/a0015108.

Schoenfeld, A. H. (2018). On replications. *Journal for Research in Mathematics Education*, *49*(1), 91–97. https://doi.org/10.5951/jresematheduc.49.1.0091.

Shayer, M., Küchemann, D. E., & Wylam, H. (1976). The distribution of Piagetian stages of thinking in British middle and secondary school children. *British Journal of Educational Psychology*, *46*(2), 164–173. https://doi.org/10.1111/j.2044-8279.1976.tb02308.x.

Steffe, L. P., & Olive, J. (2010). *Children's fractional knowledge*. Springer. https://doi.org/10.1007/978-1-4419-0591-8.

Streefland, L. (1991). *Fractions in realistic mathematics education: A paradigm of developmental research*. Springer. https://doi.org/10.1007/978-94-011-3168-1.

The jamovi project (2022). *Jamovi.* (Version 2.3) (Computer Software). https://www.jamovi.org.

Undervisningsministeriet. (1960). *Undervisningsvejledning for folkeskolen. Den blå betænkning, bind 1.* Betænkning nr. 253 afgivet af det af Undervisningsministeriet under 1. September 1958 nedsatte Læseplansudvalg (Teaching guide for primary school. The blue report, volume 1. Report no. 253 issued by the Curriculum Committee set up by the Ministry of Education on 1 September 1958). S. L. Møllers Bogtrykkeri. https://www.betænkninger.dk/wp-content/uploads/2021/02/253.pdf.