Using TIMSS data to explore Nordic struggles with fractions

Anne-Catherine W G Lehre

Master of Science

Assessment, Measurement, and Evaluation

30 credits

Centre for Educational Measurement (CEMO)

Faculty of Educational Sciences

November 2023

Popular abstract

Feeling uncertain when you deal with fractions? You are not alone! Dating back more than 4,000 years, fractions have always played a major role in society. Despite its importance, research shows that many pupils, students, and adults struggle with fractions. A blogger pointed to Nordic grade 8 students struggling more than other students around the globe when trying to solve the task: «Which shows a correct method for finding ⅓ – ¼? », taken from an international large-scale assessment in education called Trends in Mathematics and Science Study (TIMSS). Using TIMSS data from several countries and spanning almost 20 years, we explored the claim that Nordic students struggle more with fractions and/or common denominator tasks than students elsewhere in the world. Focusing on tasks in mathematics, the results show that Nordic students, neither struggle more on fractions tasks than non-fraction tasks, nor struggle more on common denominator tasks than non-common denominator tasks. A similar pattern is seen within each of the countries included in this study. However, not limited to the specific student cohort mentioned in the blog, but also other cohorts of Nordic students struggle with specific tasks similar to «Which shows a correct method for finding ⅓ – ¼? ». Future research could explore why: For example, are there noticeable cross-cultural differences in instructions and textbooks used.

Acknowledgements

The present study has been carried out at the Faculty of Educational Sciences, University of Oslo, Norway. Thank you to staff and faculty at CEMO for their passion for the master program and for their patience, true dedication, and concern for their students.

My supervisor, Johan Braeken, deserves my deepest gratitude for fruitful guidance, encouragement, interesting discussions, and invaluable support. Thank you for putting up with me during frustrations of R coding, not enough hours in the day, and writer's block, and for sharing your statistical and R knowledge. It has been a pleasure working with you.

Thank you to fellow students for creating such a friendly, encouraging, and interesting environment. We have shared steep learning curves, great discussions, laughs, and stressful times.

Also thank you to Tony Tan for helping with R coding, and to Hege Kaarstein for reliability coding of items and for fruitful discussions concerning the TIMSS data.

I dedicate this thesis to my parents for showing me the value of education, to my friends and my family for their patience, and to my children for their unconditional love.

Oslo, November 2023

Abstract

Proficiency with fractions is an important foundation for learning probability and more advanced mathematics, and for participation in various occupations. Research shows that understanding fractions is a struggle for many students. A blog discussion (http://akarlin.com/2013/07/scandinavians-cant-do-fractions) pointed to one specific item (item M052228: «Which shows a correct method for finding ⅓ – ¼? ») in the International Association for the Evaluation of Educational Achievement's (IEA) 2011 Trends in International Mathematics and Science Study (TIMSS) where grade 8 students from the Nordic countries struggle more than most students around the globe. To shed some light on the generalizability of this phenomenon, data from several TIMSS cycles was used in an explanatory item response theory approach to investigate the performance of grade 8 students in Nordic versus non-Nordic countries on items involving fractions and on fraction items requiring a common denominator. The results show that not limited to the specific student cohort mentioned in the blog, but also other cohorts, Nordic students show inferior performance on specific items similar to «Which shows a correct method for finding ⅓ – ¼? ». However, problems with this specific type of fraction items cannot be attributed to a general problem with common denominator items, nor with fraction items as a group. If the Nordic students' inferior performance on this specific type of fraction items is not a generalizable issue with fractions or common denominators, then future research could explore why: For example, are there noticeable cross-cultural differences in textbooks used and instructions.

Keywords: fractions; TIMSS; explanatory item response modeling; item difficulty

Using TIMSS data to explore Nordic struggles with common fractions

Fractions, together with decimals, are common in fields such as mathematics, economics, science, and psychology (Haryanto, 2019). Proficiency with fractions is fundamental for everyday quantitative reasoning (Yeo & Webel, 2022), and can influence success in many professions (Lortie-Forgues et al., 2015). A strong understanding of fractions is crucial for overall mathematical understanding (Mou et al., 2016; Siegler et al., 2011; Torbeyns et al., 2015), and a key factor underlying students' general mathematics achievement (Torbeyns et al., 2015). Proficiency with fractions is an important foundation for advanced mathematical content areas such as learning probability, algebra, and more advanced mathematics (Torbeyns et al., 2015). Fraction knowledge predicts mathematics achievement of elementary pupils in high school, even after controlling for other numerical knowledge, working memory, and intelligence (Bailey et al., 2014; Booth et al., 2014; Booth & Newton, 2012; DeWolf et al., 2015; Hurst & Cordes, 2018; Siegler et al., 2012). Children's reasoning about fractions relates to their algebraic reasoning (e.g., Empson et al., 2011; Hackenberg, 2013; Tunc Pekkan, 2008).

Despite its importance, research shows that difficulties in understanding fractions and their operations applies for younger students (Ni & Zhou, 2005; Vamvakoussi, 2015), secondary school students (Moss & Case, 1999; Siegler & Pyke, 2013), University students (Chinnappan & Forrester, 2014; Hanson & Hogan, 2000), and adults (DeWolf & Vosniadou, 2015). The concept of fractions is said to be the most challenging mathematical concept to learn (National Mathematics Advisory Panel, 2008). It is well documented that students struggle with learning fractions (e.g., Durkin & Rittle-Johnson, 2015; Ni & Zhou, 2005; Siegler et al., 2011; Stafylidou & Vosniadou, 2004; Vamvakoussi, 2015; Van Dooren et al., 2015), and that understanding even common fractions is difficult for children (Ni & Zhou, 2005; Vamvakoussi, 2015). Several factors have been identified as contributing to students' difficulties in learning fractions (Charalambous & Pitta-Pantazi, 2007), and it is generally agreed that the complexity of learning fractions may lie in the fact that fractions is a multifaceted construct (Haryanto, 2019; Kieren, 1976, 1993; Lamon, 2006).

In Norway, teachers at secondary schools and upper secondary schools have reported that students struggle with fraction tasks (Lande, 2022). A blogger (http://akarlin.com/2013/07/scandinavians-cant-do-fractions; Appendix III) commented on how Nordic students showed deficient performance on one fraction item («Which shows a correct method for finding ⅓ – ¼? ») from the 2011 Trends in Mathematics and Science Study (TIMSS). A master thesis shows that Norwegian grade 8 students showed deficit performance on four out of six selected fraction items from TIMSS 2003 (Utgård, 2008). The item showing the largest deficit performance for Norwegian students involved the concept of common denominator (Utgård, 2008), just like the item identified by the blogger. Herein, we will investigate whether the deficient performance on common denominator items, as identified by the blogger and the master student, extends to other comparable items in TIMSS, whether fractions and probability in general, or common denominator, is key to elicit deficient performance. The focus will be on the difficulty differences between items.

Theoretical Framework

Item Variance

TIMSS is one of several international large-scale assessments (ILSAs) that collect performance and background variables from educational systems (Rocher & Hastedt, 2020). The performance profiles of each participating country in ILSAs provide basic indicators of student performance related to demographic, socio-economic, and educational indicators (Mullis & Martin, 2017). Commonly ILSAs report average country achievement scores at the level of the subject (e.g. mathematics) or within various broadly defined content and cognitive domains (Mullis et al., 2020). In TIMSS, for example, the mathematics' construct is organized as a two dimensional matrix where the content domain is based on Tyler's (1949) categorization framework for topics, and the cognitive domain is based on Bloom's taxonomy (Bloom, 1956). For grade 8 students participating in TIMSS, average achievement results in mathematics are reported for the content domains of *number*, *algebra*, *geometry and measurement*, and *data and probability* and for the cognitive domains of *knowing*, *applying*, and *reasoning* (Mullis et al., 2020; Philpot et al., 2023). However, the reports have limited value-added for teachers and curriculum designers (Daus et al., 2019) as the performance profiles lack «information about countries' relative strengths and weaknesses regarding different items or topics» (Marcq & Braeken, 2022, p. 346). Investigating the within country response variation of items may be one way to uncover potential information (Marcq & Braeken, 2022), and to produce value-added information for curriculum designers, test developers, and teachers by providing more detailed performance profiles of each country (Daus et al., 2019).

An example of one study exploring the item-side in an ILSA, is a study by Daus et al. (2019) that explored within country response variation of science items in TIMSS 2011. The study provided a finer-grained analysis of Norwegian grade 8 students' strengths and weaknesses in specific science domains and topics. Topics that were relative harder or relative easier compared to other topics for Norwegian students were identified (Daus et al., 2019). In addition, the Norwegian students' relative strength and weakness profile was compared with the profile of the average TIMSS participating countries. The Norwegian country profile showed that students had a higher proportion correct responses to items concerning Earth Science, compared both internally and externally, but that Physics was the most difficult domain (Daus et al., 2019). Such profiles may provide teachers with valuable insight into what topics their students may find difficult, and thereby adjust their teaching accordingly.

Using TIMSS 2011 data, Daus et al. (2019) found that «the variance components of the hierarchical item response model showed that about 30% of the variation in responses was due to the item characteristics, compared with only 15% due to the pupil abilities». This implies that, the item mattered more than which student responded to it. This is in agreement with a study using PISA 2018 data to estimate variance components of schools, pupils, and items, showing that «which items were responded to by a student mattered more than which pupils responded to the items» (Marcq & Braeken, 2022). Quantifying the item variance, the current study will investigate TIMSS mathematics' items and explore the claim that Nordic students show deficit performance on a specific type of fraction item, by providing more detailed within country mathematics performance profiles of herein selected countries.

Fractions

The first written fractions appear in the Babylonian and Egyptian cultures some 4000 years ago (Cajori, 1894; Kieren, 1976). While the Egyptians used a duodecimal system expressing rational numbers in terms of unit fractions, the Babylonians used a sexagesimal system where, for example, ½ was designated as 30 (reader was expected to supply the word «sixtieth»). The astronomer Ptolemaeus (c. 100 - 168 AD) introduced the sexagesimal system into Greece, and the sexagesimal fraction were used in mathematics and astronomy throughout Europe up to the sixteenth century (Cajori, 1894). In 1858, decimal fractions were introduced in Europe by Simon Stevin (1548-1620) in his book *De Thiende* (Cajori, 1894), but decimal fractions were first developed and used in China, tracing back to the fourteenth century BC (Merzbach & Boyer, 2011).

Stevin was the first to define a fractional number as «a part of the parts of a whole number» (Cajori, 1894; Park et al., 2013). To distinguish fractions used in mathematics from the sexagesimal fraction used in astronomy, the term simple/common/vulgar fraction (herein:

«common fraction») was coined. The common fraction (from Latin: *fractus,* «broken») represents the part-whole meaning of rational numbers, expressed in the «form a/b, where a, b are integers/whole numbers and b ≠ 0» (Kieren, 1976; Ni & Zhou, 2005). Fractions have distinct symbolic notation and properties, are unbounded-infinite numbers, and are a complex construct belonging to the category rational numbers (Behr et al., 1993; Kieren, 1976; Stafylidou & Vosniadou, 2004). Rational numbers have been described as a «mega-construct» (Kieren, 1980) including multiple sub-constructs (Behr et al., 1983; Charalambous & Pitta-Pantazi, 2007; Kieren, 1976). These sub-constructs are related, and they include *part-whole* comparisons (equal parts of a larger unit), *quotients* (division), *ratios* (comparing two entities), *operators* (transform a quantity into another quantity with a smaller or bigger value), or *measures* (measurement point) on a number line (Table 1).

Table 1

Fraction sub-constructs (Kieren 1976, 1980): exemplified by meaning of «*2/5*»

Sub-construct	Example
Part-whole	2 slices out of a whole cake with 5 slices
Quotient	2 divided by 5
Ratio	2 oranges: 5 apples
Operator	Scaling factor
Measure	The distance that is 2/5 of the way between 0 and 1

Each of Kieren's sub-constructs capture different aspects of the concept of fractions (Table 1), and for students to develop robust fractions knowledge, they must understand and integrate these partially overlapping sub-constructs (Behr et al., 1993). The complexity also involves different arithmetic procedures depending on the operations (for example: addition versus division of fractions), the equality of the denominators, and the type of fraction (Braithwaite et al., 2017; Lortie-Forgues et al., 2015; Siegler & Pyke, 2013). A fraction may for example be classified as proper or improper depending on the values of the numerator and the denominator, mixed-, equivalent-, unit - or decimal fraction, percentage, or ratio. In short, the concept of fractions captures multiple ideas through various notations (Bobis, 2011).

Common Denominator

The use of common denominators are traced back to the Chinese and the fourteenth century BC (Merzbach & Boyer, 2011). Common fractions have common denominator when the denominator of two or more fractions are equal (*Vocabulary.Com.*, n.d.). Before adding and subtracting fractions, the fractions must share a common denominator (*Vocabulary.Com.*, n.d.). The observation that students may struggle with common denominators in adding and subtracting fractions is not a new phenomenon. Already in the 1930s, educators in the United States launched a study to decide when (which grade) students were mature enough to master addition and subtraction of fractions (Charalambous et al., 2010; Raths, 1932). Explorations of why students make errors in adding or subtracting fractions, has been extensively studied, and points to several cognitive factors (Charalambous et al., 2010). One such factor is that students may believe that whole-number arithmetic may be valid for fractions (Gabriel et al., 2013; Siegler et al., 2011). Struggles with addition and subtraction with fractions cannot be separated from struggles with fractions in general (Charalambous et al., 2010).

A common denominator is also used in the Eudoxian definition of equality of ratios, a process similar to the cross-multiplication of fractions: «$a/b = c/d$ if and only if given integers m and n, whenever $ma<nb$, then $mc<nd$, or if $ma=nb$, then $mc=nd$, or if $ma>nb$, then $mc>nd$» (Merzbach & Boyer, 2011, p. 80). According to Fish (1874) «the ratio of two fractions is obtained by reducing them to a common denominator, when they are to each other as their numerators» (Fish, 1874, p. 388). Common denominators are also needed when comparing the magnitude of various fractions with different denominators and/or decimal numbers.

Fraction Learning Progression – Student Issues

Rational numbers and fractions are notoriously difficult for students to master, and often interferences from natural number knowledge, and student interpretation of rational number notation, cause issues (Ni & Zhou, 2005). Students tend to believe that a fraction is an arrangement of two parts, not a single quantity (Yeo & Webel, 2022), and this natural number bias often lead students to make errors when dealing with fractions and decimals (Lamon, 2006). In addition, the direction of effects in whole number arithmetic may cause bias when dealing with multiplication and division of fractions smaller than one (Mostert & Hickendorff, 2023; Siegler & Lortie-Forgues, 2015). Different operations with fractions require different arithmetic procedures making solving fractions problems more difficult and abstract for students (Braithwaite et al., 2017; Lortie-Forgues et al., 2015; Mostert & Hickendorff, 2023).

Attempting to explain students' knowledge and understanding of fractions, researchers have developed frameworks to estimate students' sophistication levels (Hachkenberg et al., 2016; Yeo & Webel, 2022). The theoretical framework of Kieren (Kieren, 1976, 1980) forms the basis of existing frameworks for fraction learning progression. Kieren's five proposed sub-constructs in the interpretation of fractions (Table 1), was, for instance, adopted by Arieli-Attali and Cayton-Hodges (2014) as central concepts and principles of their «rational number learning progression» model. In addition to Kieren's sub-constructs, other central concepts and principles as half and halving procedures, unit fraction, decimals, place value, and equivalent fractions were included, and used to construct progress variables. These progress variables form the basis of a structured six-level progression of students in rational number learning model: prior knowledge (half and halving), early part-whole understanding, fraction as unit, fraction as single number and fraction as measure, representational fluency, and a general model of a rational number (Arieli-Attali & Cayton-Hodges, 2014; Haryanto, 2019).

The learning progression models above does not distinguish between the two essential knowledge dimensions in mathematics learning, conceptual and procedural knowledge (Hibert & Lefevre, 1986; Rittle-Johnson & Schneider, 2014). However, fractions competencies can be divided into conceptual and procedural knowledge (Bailey et al., 2014). Conceptual knowledge refers to «the understanding of the nature and mathematical properties of fractions, including that a fraction consists of a numerator and a denominator, its magnitude is determined by the numerator-denominator relation, and that these magnitudes can be ordered on a number line» (Mou et al., 2016, p. 141). In other words, conceptual understanding refers to the understanding of the principles or «knowing why» (Baroody, 2003) whereas procedural knowledge refers to the understanding of how to solve fraction problems (Bailey et al., 2014). Both competencies are essential for the understanding of fractions (Siegler et al., 2011), and many tasks require students to use both conceptual and procedural strategies to solve the tasks (Faulkenberry, 2013). One model of fraction learning progression that includes both conceptual and procedural knowledge was proposed by Haryanto (2019). In his framework, he describes five levels of student knowledge/proficiency with fractions, from no knowledge (level 1) to fluency (level 5), on five conceptual sub-constructs (part-whole, measure, infinity, additive structure, and multiplicative structure) and two procedural constructs (additive- and multiplicative operations).

Students' difficulties in learning fractions have been explained by fractions symbolic notation and properties (Stafylidou & Vosniadou, 2004), the complexity of the concept (Behr et al., 1993; Kieren, 1976), and natural number bias (Lamon, 2006). In addition, student performance on fractions is also influenced by their mathematics' teachers' content knowledge (Bobis, 2011), the textbooks used (Charalambous et al., 2010), and cross-cultural differences in instruction and achievement (Charalambous et al., 2010; Klacznski et al., 2019).

The Present Study

The present study investigates whether the deficient performance on the one item identified by the blog http://akarlin.com/2013/07/scandinavians-cant-do-fractions extends to other comparable items in TIMSS, whether fractions and probability in general, or common denominator, is key to elicit deficient performance. Taking both a between country perspective, and a within country perspective (generic, fraction computation, or common denominator), the focus will be on the difficulty differences between items. Given that the focus will be on the items and item variance, the study will take an exploratory approach (Tukey, 1977). The main research question is:

RQ: *Are specific fraction items more difficult for Nordic students than other mathematics items compared to students in other countries?*

The following core research objectives (ROs) will be addressed:

RO1: *Response option analysis of the item and its clones*

The objective is to investigate descriptive proportions of chosen response alternatives on (i) item M052228: «Which shows a correct method for finding ⅓ – ¼? » and (ii) its clones across TIMSS cycles (from 2003 to 2019), by Nordic countries contrasted to international averages and reference countries.

RO2: *Item Response Modelling: Difficulty ~ Item Features*

The objective is to quantify the relative magnitude of response variation due to differences in item difficulties in TIMSS 2019. The focus will be on the items and whether fraction-related items are more difficult than non-fraction items in mathematics, and whether common denominator items are more difficult than other items in mathematics, for Nordic students. Nordic students will be contrasted to reference countries.

Method

TIMSS is an international cross-sectional survey, developed by the International Association for the Evaluation of Educational Achievement (IEA), measuring grade 4 and grade 8 pupils' competence in mathematics and science, in addition to mapping factors that promote learning (Mullis et al., 2020; Mullis & Martin, 2013). Every 4th year, a representative sample of pupils is selected for participation in the survey (LaRoche et al., 2020). The selection of pupils follows a two-stage stratified design in each participating country. First, the schools are sampled with probability proportional to their size, and then 1-2 intact classes are drawn from the targeted grade for each of the participating schools. The proportion of boys and girls tends to be evenly distributed (Fishbein et al., 2021). In Norway, the selection of pupils comprises approx. 7% of the population of the targeted grades (Kaarstein et al., 2020).

Focus item: Response Option Analysis across Cycles

For research objective one, the descriptive, weighted proportions of chosen response alternatives to item M052228 and its clones were investigated, and the outcome measure is a percentile rank score for the international country average.

Items

To find the focus item and its clones, grade 8 data on scored item responses in mathematics from TIMSS 2003, TIMSS 2007, TIMSS 2011, TIMSS 2015, and TIMSS 2019, were investigated. Among the roughly 200 mathematics items in a TIMSS cycle, we found one item clone in each of the following cycles: TIMSS 2003, TIMSS 2007 and TIMSS 2019.

In addition to the focus item M052228 from TIMSS 2011 (Appendix III; Figure S1) discussed in the blogpost (Appendix III; blog), a similar item clone, M032416 (Appendix III; Figure S2), used in TIMSS 2003 and again in TIMSS 2007, was investigated. In TIMSS 2019, another clone appeared, ME72038 («Which shows a correct method for finding $\frac{1}{5} - \frac{1}{8}$? »), but no data were provided by IEA due to «severe differential item functioning» (Martin et al., 2020, p. 10_63).

The item M052228 («Which shows a correct method for finding $\frac{1}{3} - \frac{1}{4}$? ») and its clone M032416 («Which shows a correct method for finding $\frac{1}{5} - \frac{1}{3}$? ») are multiple choice items with four response options each, A-D. The four response options have the same structure, in the same order, on both items (Appendix III; Figures S1-S2). Response option A has 1-1 as numerator, and the denominator is the smaller denominator subtracted from the larger denominator of the two fractions in the question. Response option B is similar to response option A, the only difference being that the numerator is set to 1. In response option C, the numerator is the second denominator subtracted from the first denominator, and the product of the two denominators from the question is the denominator. Response option D is the correct response option on both items.

Sample

The Nordic countries (Norway, Sweden, and Finland) will be compared to one Baltic country (Lithuania), two European countries (England and Italy), and two non-Nordic countries with similar overall scores in mathematics (Cyprus and Romania). In addition, two high scoring countries (Korea and Singapore), two low scoring countries (Botswana and South Africa), and the international average, are included as points of references.

For the focus item, M052228 in TIMSS 2011, the number of within country student responses range from 538 (Norway) to 1714 (South Africa). For its clone, M032416, the number of students responses were ranging from 230 (England) to 746 (South Africa) in TIMSS 2003 and from 561 (England) to 739 (Sweden) in TIMSS 2007. The exact numbers of students responding to the focus item and its clones within each country, are presented in Table 2 in the result section.

Statistical Analyses

To descriptively compare the performance on the focus item and the overall average mathematics' results across the countries, the official TIMSS numbers of proportion correct student responses (weighted) and the average mathematics' score for each country were collected from the official TIMSS Almanacks of each cycle (Mullis et al., 2012). The Almanacks can be found via the TIMSS official webpage: https://timssandpirls.bc.edu/timss-landing.html.

The overall mathematics' score for each country has been transformed to [0,1] range by transforming it to a normal (M = 500, SD=100) percentile equivalent score.

Item Difficulty as a Function of Item Features

For research objective two, an explanatory item response theory (IRT) modelling approach (De Boeck & Wilson, 2011) was used to study the relationship between the type of items and their empirical difficulty. One-parameter logistic item response models (1 PL models) were employed, providing as outcome a Rasch-like descriptive baseline model to be compared to a model where item difficulties are a function of the predictors (fraction; common denominator). Then, within-country item difficulties estimated by the descriptive baseline models were visualized as landscape plots providing direct pairwise comparions of the item difficulty ranking differences between countries (Braeken & Paap, 2020; Navarro et al., 2004).

Items

All items in mathematics from TIMSS 2019 were included in the categorization (n = 211). Coding according to the item features below, showed that 68 of the 211 items in mathematics were categorized as fraction items. Of the 68 fraction items, 38 were categorized as common denominator items.

The TIMSS 2019 items ME62048, ME62342, ME62345B, ME72038, and ME72211B were deleted from the sample since no data were provided by IEA due to either «poor discrimination» and «severe differential item functioning» (Martin et al., 2020, p. 10_63).

Item features. As a basis for the categorization, a classic categorization theory was used. A mathematics item had to meet the defined criteria of a category to belong into the category. The focus herein was on fractions and common denominator items. Thus, the mathematics items were binary coded (0/1, where 1 meant the item was included into the category) for inclusion into the following categories (Figure 1):

1. Mathematics
2. Fractions
3. Common denominator

All mathematics items from TIMSS were included in the first category *Mathematics.* The second categorization separates items into either a fraction item or a non-fraction item. Items belonging to the *Fraction* category were defined broadly due to the complexity of the construct (for example: Behr et al., 1993; Kieren, 1976; Stafylidou & Vosniadou, 2004). This includeed Kieren's sub-constructs (1976; 1980): part-whole comparisons, quotients, ratios, operators, and measures on a number line. A fraction item could also ask for or be dependent on:

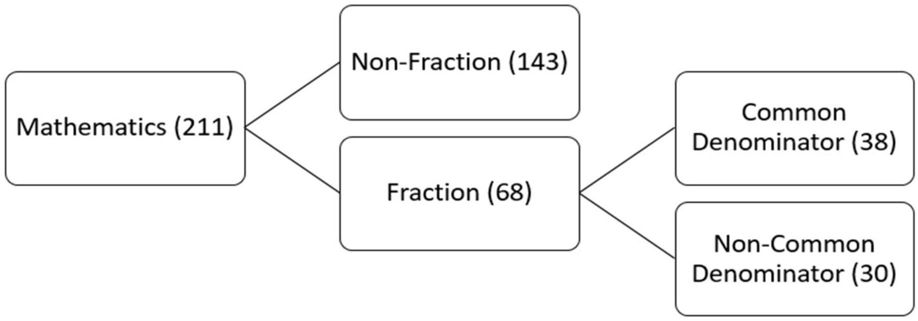
- using rules for adding, subtracting, or multiplying fractions (also in combination with whole numbers) or explaining how to
- finding common denominator for fractions or explaining how to
- finding and/or handling equivalent fractions (e.g., expanding, cancelling, or comparing) or explaining how to
- using fractions to solve proportional situations or explaining how to

The third categorization separated fraction items that involved the use of a common denominator, including finding common denominator for fractions or explaining how to use or find a common denominator, from fraction items that did not require a common denominator.

The categories are nested so that all items belonged to the category *Mathematics.* All the items were then placed in a second layer category, either *Fraction* or *Non-Fraction*. All the fraction items were in turn placed into a third layer category, either *Non-Common-Denominator* or *Common-Denominator* (Figure 1).

Figure 1

Overview of item features: the nested categories of the item features



Note: Items were categorized into categories from left (*Mathematics*) to right (*(Non-)Common Denominator*).

Inter-rater Reliability. All the mathematics items across all the cycles of TIMSS (n =1,064 items) were coded by the author according to the framework described above. To check the rater reliability of the coding of the items, the TIMSS Norwegian National Research Coordinator, that holds a PhD in mathematics didactics, coded items from TIMSS 2019 (n = 211) according to this study's item features. The agreement between the two coders was 94.4 %, Cohen's κ = 0.87 (Cohen, 1960). According to Cohen, if a κ is between 0.81 and 1.00, the agreement is considered as «almost perfect agreement» (Cohen, 1960).

Sample

For the second research objectives, mathematics items from the most recent cycle of TIMSS were investigated. Only the selected countries explored for research objective one, that were participating digitally, were included: Norway, Sweden, Finland, England, Italy, Lithuania, Singapore, and Korea. Romania, South Africa, and Botswana participated in the paper-based version of TIMSS 2019, and further analyses for these countries were not undertaken.

Students responding correctly to 5 or less items were deleted, resulting in deletion of 279 students out of 61,751 students (0.45 %).

Statistical Analyses

Basic data management procedures were done in the statistical software environment R (R Core Team, 2021). TIMSS student responses on items were recoded into binary codes (0/1), where the code 0 was used for an incorrect response, and the code 1 was used for a correct response. Correct student responses to partial credit items were recoded into 1. All missing data, whether omitted or not reached, were recoded into NA. Details on the data management and the recoding are to be found in Appendix II.

The assessment pool of items are distributed into sets of 14 booklets by means of a matrix sampling approach, and each student complete one booklet only (Mullis & Martin, 2017). In the grade 8 mathematics assessment, 24 – 32 items are given to each student (Mullis & Martin, 2017). The number of students responding to items in a booklet in TIMSS 2019 varied both within and between countries, ranging from 495 students (England, booklet 12) to 702 students (Finland, booklet 4) for the countries studied herein.

Item Response Models. Assuming equally discriminating items, the one-parameter logistic item response model (1 PL model) was chosen. In this model, all assessment items contribute equally to the latent construct (Rasch, 1980; von Davier, 2016). The model assesses the latent abilities of the test taker based on the correctness of their answers on test items with different difficulty levels. First a descriptive baseline model (null model) as modelled. Then the baseline model was extended by including item characteristics (herein: fraction and common denominator) as predictors for the item difficulty

A random-person random-item explanatory item response modelling approach (De Boeck, 2008) was formulated, inspired by the approach of Marcq and Braeken (2022). The total response variance was partitioned into variation from persons (θ_p) and variation from items (β_i), and the model for probability of a correct response (π_{pi}) is defined as:

$$\text{Logit}\,(\pi_{pi}) = \beta_0 + \theta_p + \beta_i$$

The fixed effect (β_0) corresponds to the estimated logit for the probability of a correct item response on an «average item» from TIMSS 2019 by an «average student». The random effects for persons (θ_p) and items (β_i) were assumed to follow an independent normal distribution with means equal to zero and variances σ^2_θ and σ^2_i, respectively. This model implies that the total observed response variation (σ^2_{tot}) is:

$$\sigma^2_{tot} = \sigma^2_\theta + \sigma^2_i + \pi^2/3$$

where $\pi^2/3$ is «the distribution-specific residual variance from the standard logistic distribution due to the applied link function accounting for the binary nature of a response» (Marcq & Braeken, 2022, p. 335), in other words, the error variance for the latent continuous response underlying the binary one. The IRT models were calibrated for each country separately, and it was expected that the predictors (fractions and common denominators) would make the items more difficult for the Nordic students.

Ranking Plots: Landscape Model. Item difficulties from the descriptive baseline models were estimated on country basis, and not restricted to be invariant across countries to allow each country to show its unique country profile in the mathematics domain of TIMSS 2019. Item were then ranked based on their item easiness, from difficult to easy, under the descriptive baseline model for each country. The within-country ranks were visualized as so-called landscape plots (Braeken & Paap, 2020; Navarro et al., 2004) providing direct pairwise comparions of the item easiness ranking differences between countries. The ranks make abstraction of overall differences in mathematics achievement between countries but allows for comparison of relative item easiness across countries (for example, to what extent does the set of easiest items overlap across countries). All items on or near the diagonal reflect cross-country correspondence in relative easiness of an item. The further away the plotted item is from the diagonal, the bigger the relative difference in item easiness. The landscape plots are used herein in an exploratory fashion to detect whether big deviations occur for specific fraction or common denominator items.

The analyses were performed using the R package lme4 (Bates et al., 2015) in the open-source statistical software R (R Core Team, 2021). The R code is found in Appendix II. The data are publicly available online at: (https://timssandpirls.bc.edu/databases-landing.html), and are not subject for notification to the Norwegian Centre for Research Data (NSD) under the General Data Protection Regulation (GDPR) (Appendix I).

Results

Focus item: Response Option Analysis across Cycles

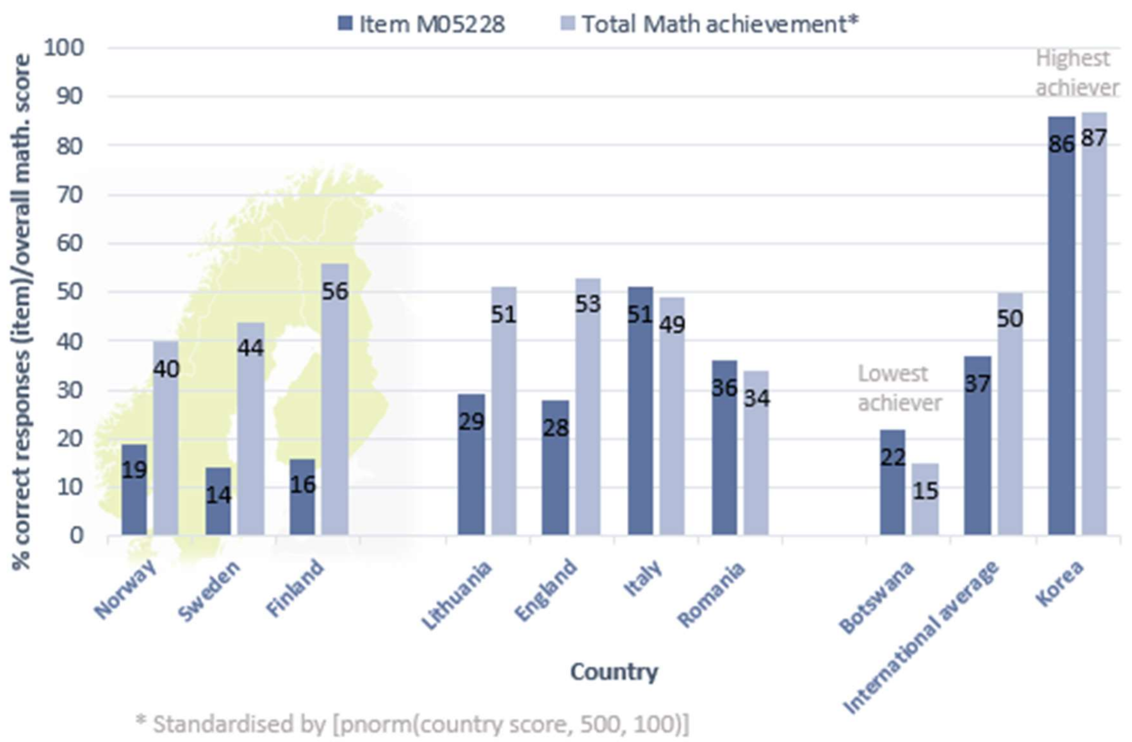
To investigate whether Nordic students struggle more on the specific item mentioned in the blogpost (item M052228), and a similar items (item M032416) across cycles in TIMSS, descriptives from several countries were compared (Appendix III, Table S2).

Item M052228 from blogpost

Percent Correct responses. Figure 2 shows the overall mathematics' score for each country transformed into a percentile equivalent score, and percent correct student responses, for the Nordic countries compared to Lithuania (Baltic country), England and Italy (European countries), Romania (similar overall score in mathematics as Nordic countries), the lowest achiever (Botswana), the international average, and the highest achiever (Korea). 14-19 % of the Nordic students gave a correct answer to the item which is lower than those of students from all the selected countries herein. The percent correct student responses for this item were between 28 % to 51 % for the European countries compared to 22 % in Botswana and 86 % in Korea. The international average was 37 % correct student responses on this item. The Nordic students show inferior performance on the specific item M052228 as mentioned in the blogpost (Appendix III).

Figure 2

Overall Mathematics Score and Percent Correct Student Responses on TIMSS 2011 Grade 8 Item M052228, across Nordic Countries Compared to Selected Countries and the International Average (Foy et al., 2013).



Note: Percent correct student responses on item M05228 («Which shows a correct method for finding ⅓ – ¼? ») for selected countries. For points of references, the overall mathematics' score for each country standardized into a percentile equivalent score

Preferred distractor. The focus item from TIMSS 2011 is a multiple-choice item with four response options (Figure 3). When investigating the within country response patterns (Appendix III, Table S2; Figure 3), the most attractive distractor within the Nordic countries and the lowest achieving countries (South Africa and Botswana) is A. Distractor B is the second most attractive choice of all these countries. The European countries have the opposite pattern, with distractor B being the most attractive distractor for the students and distractor A being the second most likely student choice. Korea and Singapore show a different pattern, with the lowest number of incorrect student responses on distractor A.

Figure 3

TIMSS Item M052228 and Most Attractive Distractor Responses for Grade 8 Student Responses for Selected Countries (Foy et al., 2013).

Which shows a correct method for finding $\frac{1}{3}-\frac{1}{4}$?

(A) $\frac{1-1}{4-3}$ Nordic; Low achieving countries

(B) $\frac{1}{4-3}$ European; Korea

(C) $\frac{3-4}{3\times 4}$ Singapore

(D) $\frac{4-3}{3\times 4}$ (Correct answer)

Note: Of students choosing a distractor over the correct answer (D), distractor A is the most attractive choice for students from the Nordic and the low achieving countries. European and Korean students were attracted to distractor B, and Singaporean students had distractor C as their most attractive choice. SOURCE item M052228: TIMSS 2011 Assessment. Copyright © 2013 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, Chestnut Hill, MA and International Association for the Evaluation of Educational Achievement (IEA), IEA Secretariat, Amsterdam, the Netherlands.

Clone of Item M052228: Item M032416

Percent correct responses. A clone (item M032416) of the blogpost item (item M052228) had been issued in TIMSS 2003 and TIMSS 2007. In general, the Nordic students showed inferior performance on the clone item through the assessment cycles (Appendix III; Table S1). Only the country with the lowest overall average mathematics score, South Africa, and England in the 2003 cycle, have as low percent correct student responses as the Nordic countries on item M032416. Cyprus and Romania have a similar overall mathematics score as Norway, but both these countries have a higher percentage of correct student responses on the focus item, ranging from 15 % (Cyprus, TIMSS 2003) to 39.7 % (Romania, TIMSS 2003) (Appendix III; Table S1). The range of correct student responses on the clone item for the selected European countries herein is from 10.6 % (England, TIMSS 2003) to 47.0 % (Italy, TIMSS 2007), with an average of 29 % correct student responses. In the highest achieving countries, Korea and Singapore, the average percent correct student responses across cycles for item M032416 is 75.3 % (range: 69.8 % - 79.2 %).

Preferred distractor. When investigating the within country response patterns for the clone item across cycles (Appendix III, Table S2), the Norwegian students found distractor A the most attractive, and distractor B the second most attractive, in both TIMSS cycles. The same pattern was shared with England in TIMSS 2003 and with Sweden in TIMSS 2007. Students from the other countries tended to pick distractor B over distractor A, except from students from Singapore and Korea that showed a different pattern from the other investigated countries (Appendix III, Table S2).

To sum up, the Norwegian students chose the same distractor (A) across all the cycles. The Swedish students also chose the same distractor (A) as the Norwegian students in all cycles, except in TIMSS 2003. The Finish students also favored distractor A in TIMSS 2011, but they did not participate in the other cycles of TIMSS. Students from the other European countries and the Asian countries did not share the same attraction to distractor A as the Nordic students did across the TIMSS cycles.

Item Difficulty as a Function of Item Features

To test for generalizability of the claim that Nordic students show deficit performance on either common denominator items or fraction items in general, a cross-classified mixed model was run for the selected countries participating in TIMSS 2019. Table 2 reports the estimated parameters of item difficulty for all the items in mathematics for each country: The null-models (baseline) and the models where item features (fractions and common denominators) are predictors for item difficulty.

Table 2

Parameter Estimates of the Cross-Classified Mixed Effect Model for the Selected Countries from TIMSS 2019

	Norway	Sweden	Finland	England	Italy	Lithuania	Singapore	Korea

N_p	4,517	3,959	4,826	3,338	3,617	3,822	4,842	3,855
N_i	204	204	204	204	204	204	204	204
N_r	116,603	103,811	132,075	89,720	96,258	104,743	139,775	110,463
M_0	Norway	Sweden	Finland	England	Italy	Lithuania	Singapore	Korea
β_0	-0.38	-0.41	-0.47	-0.27	-0.54	-0.32	1.13	1.09
(SE)	(0.08)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)
$\sigma^2{}_p$	1.31	1.23	1.18	1.50	1.04	1.51	2.17	2.46
% $_p$	22.2	21.1	19.9	25.0	19.0	25.3	33.7	35.7
$\sigma^2{}_i$	1.30	1.30	1.47	1.20	1.12	1.17	0.98	1.14
% $_i$	22.0	22.3	24.8	20.0	20.6	19.6	15.2	16.5
AIC	127,766	114,127	143,065	98,328	107,858	114,689	134,068	105,041
BIC	127,795	114,156	143,095	98,356	107,886	114,718	134,097	105,070
M_1	Norway	Sweden	Finland	England	Italy	Lithuania	Singapore	Korea
β_1	-0.44	-0.45	-0.53	-0.27	-0.51	-0.28	1.05	1.01
(SE)	(0.10)	(0.10)	(0.11)	(0.10)	(0.09)	(0.10)	(0.09)	(0.10)
β_{1Frac}	0.11	0.05	0.11	-0.02	-0.09	-0.13	0.23	0.24
(SE)	(0.18)	(0.18)	(0.19)	(0.18)	(0.17)	(0.17)	(0.16)	(0.17)
β_{1CoD}	0.34	0.39	0.32	0.23	-0.04	0.14	0.21	-0.04
(SE)	(0.36)	(0.36)	(0.38)	(0.35)	(0.34)	(0.34)	(0.31)	(0.34)
$\sigma^2{}_p$	1.31	1.23	1.19	1.50	1.04	1.51	2.17	2.46
$\sigma^2{}_i$	1.29	1.29	1.46	1.20	1.21	1.17	0.96	1.13
AIC	127,768	114,129	143,068	98,331	107,862	114,692	134,068	105,043
BIC	127,816	114,177	143,117	98,378	107,909	114,740	134,117	105,091
χ^2 (df=2)	1.847	1.673	1.525	0.443	0.408	0.616	3.609	2.167
p-value χ2	0.397	0.433	0.467	0.802	0.815	0.735	0.165	0.339
R^2 (fixed)	0.002	0.002	0.002	0.000	0.000	0.001	0.002	0.002

Note. The number of students, items, and responses used in the analyses are denoted as N_p, N_i, and N_r, respectively. M_0 represents the null model (baseline for all items), and M_1 represents model where item features (fraction and common denominator) are predictors for item difficulty. The estimated probability of a correct response of an average student on an average item, a non-fraction item, a fraction item, and a common denominator item, are denoted β_0, β_1, β_{1Frac}, and β_{1CoD}, respectively, and their standard errors are denoted SE. $\sigma^2{}_p$ and $\sigma^2{}_i$ are the variances of the random student and item effect, respectively. Estimates of response variance attributed to the student, % $_p$ and to the item, % $_i$. Model comparison indices are AIC, BIC, and χ^2 including its p-value. R^2 (fixed) is how much of the item variation the item predictors account for.

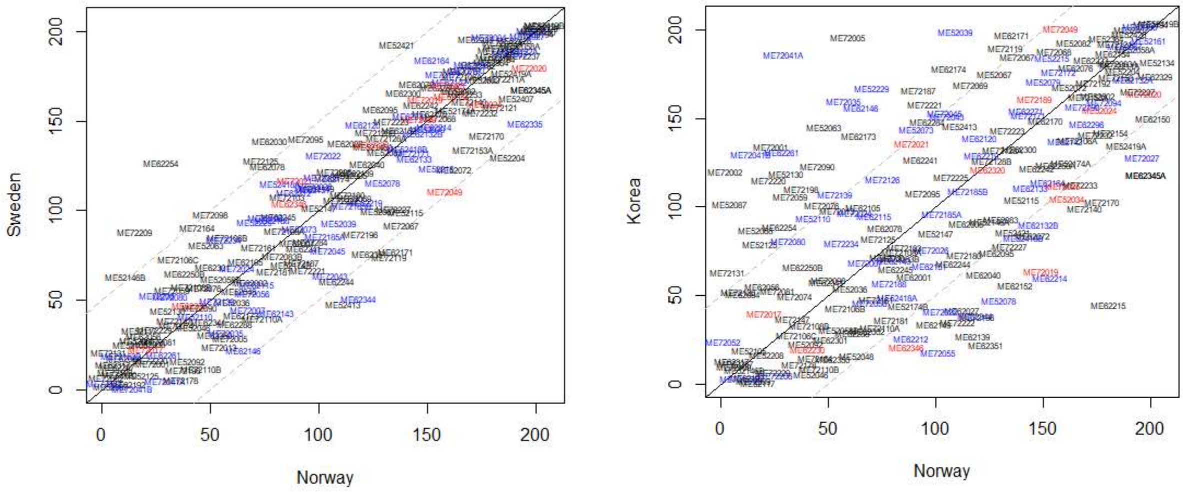
The descriptive null models for the Nordic countries show that about 20 % of response variance are attributed to individual differences in ability between persons (Table 2). For Norway and Sweden some 20 % of response variance are attributed to differences in difficulty between items, while in Finland some 25 % of response variance are attributed to differences in difficulty between items. The response variance among students from Italy shows a similar trend to Norway and Sweden, since the response variance attributed to individual differences in ability between persons, 19 %, is of a similar magnitude as the response variance attributed to differences in difficulty between items, 21 %. For England and Lithuania, the response variance attributed to individual differences in ability between persons (25.0 %) are slightly higher than the response variance attributed to differences in difficulty between items (20 %). In Singapore and Korea, the magnitudes of the response variance attributed to individual differences in ability between persons are double those attributed to differences in difficulty between items (Table 2).

The models where item features (fraction and common denominator) are predictors for item difficulty are not significantly different from the null models. The likelihood ratio tests, range from χ^2 (2) = 3.609, p-value = 0.165, for Singapore, to χ^2 (2) = 0.408, p-value = 0.815, for Italy, with the Nordic countries in the middle of the range (Table 2). The information criteria, AIC and BIC, have similar magnitudes for the null models and the item feature explanatory models for each country. The item predictor (R^2 fixed) accounts for 0.1 – 0.2 % of the item variation. The pattern is the same for all the countries included, across Nordic and non-Nordic, and independent of whether it is a high – or low scoring country.

The landscape plots in Figure 4 provide direct pairwise comparions of the item easiness ranking differences between countries. The landscape plot to the left show a comparison between Norway and an other Nordic country, Sweden. The plot of item easiness ranking differences between Norway and Sweden (Figure 4; left), shows that most items are on or near the diagonal, reflecting cross-country correspondence in relative easiness of the items. The color codes (items in mathematics are grey, fraction items are blue, and common denominator items are red) show that no blue (fraction) nor red (common denominator) items deviates in either country. The landscape plot to the left shows a comparison between Norway and a high achieving country, Korea (Figure 4). The plot of item easiness ranking differences between Norway and Korea (Figure 4; right), shows a more spread-out pattern, reflecting a lower cross-country correspondence in relative easiness of the items.

Figure 4

Landscape Plots Providing Pairwise Comparions to Check Item Easiness Ranking Differences between Countries (Herein, Norway Compared to a Nordic Country, Sweden (left), and Norway Compared to a High Achieving Country, Korea (right))



Note: The scatterplot represents item easiness for the same sample of items under an item selection rule for Norway (horistontal axis) and an item selection rule for an other country (vertical axis); Sweden (left) and Korea (right). Each code represents the pairwise combination of item easiness for one specific item. The items are color-coded so that items in mathematics are grey, fraction items are blue, and common denominator items are red. The solid black diagonal line with intercept 0 and slope 1 divides the plot: Codes on the line have the same item easiness for both countries. The further away an item is from the diagonal, the larger the difference in relative easiness ranking of the item between the two countries compared. For reference, the dashed grey diagonal lines indicate +/- 50 ranks different. The landscape plot to the left is reflecting a similar cross-country correspondence in relative easiness of the items. The landscape plot to the right is reflecting a lower cross-country correspondence in relative easiness of the items.

The landscape plots for the other selected countries are to be found in Appendix III (Figures S3-S7) alongside estimates of easiness and standard error of the easiness estimates for all the TIMSS 2019 items for the selected countries (Table S3 and Table S4). The models are country-specific and not restricted to be invariant across countries allowing each country to show its unique profile of TIMSS 2019 mathematics domain. None of the direct pairwise comparions of the item easiness ranking differences between countries show a systematic pattern within a country for either fraction or common denominator items, but there are some regional themes. The Nordic countries have the most similar item easiness rankings. Then Norway compared to other European countries show a slightly more spread-out pattern than the ones between Nordic countries. The largest differences in item easiness rankings are between the Nordic and the Asian countries (Figure 4; Figures S3-S7).

Discussion

Fractions have played an important role in society for some 4000 years (Cajori, 1894; Merzbach & Boyer, 2011), yet, research show that that difficulties in understanding fractions and their operations applies to younger students (Ni & Zhou, 2005; Vamvakoussi, 2015), University students (Chinnappan & Forrester, 2014; Hanson & Hogan, 2000), and adults (DeWolf & Vosniadou, 2015). Using TIMSS data from several countries and spanning almost 20 years, this study explored a claim that Nordic students struggle more with fractions and/or common denominator tasks than most students across the globe.

Focus item: Response Option Analysis across Cycles

Taking the fraction item «Which shows a correct method for finding ⅓ – ¼? » from TIMSS 2011 as a starting point, the results show that Nordic students show inferior performance on this specific item, as mentioned in a blogpost (Appendix III). Less than two in ten Nordic students delivered a correct answer (from 14.4 % in Sweden to 19 % in Norway), while almost four in ten students answered correctly internationally (Figure 2; Table S1). In the highest achieving countries, more than eight out of ten students gave the correct answer to this item. On a similar item issued in two consecutive cycles of TIMSS, TIMSS 2003 and TIMSS 2007, less than one in ten Norwegian and Swedish students answered correctly (Figure 2; Table S1). This is a lower percentage of correct answers on this fraction item, than for example students from Botswana, one of the lowest achieving countries in TIMSS mathematics. In Botswana, 17.5 % and 19 % of the student replies were correct on this item for TIMSS 2003 and TIMSS 2007, respectively. Comparing with other European countries, the Nordic students also show inferior performance (Table S1). In TIMSS 2019, a similar item was issued, but unfortunately no data for this item is available due to «severe differential item functioning» (Martin et al., 2020, p. 10_63). Since students from the Nordic countries show inferior performance of this type of fraction items, it would have been interesting if the data had shown that the «severe differential item functioning» was caused by the responses from Nordic studens. However, that is not possible to check as no data are published on the item clone from TIMSS 2019.

Investigating the patterns of incorrect student responses within each country, shows that students in different counties are attracted to different distractors (Figure 3; Table S2). The Norwegian students are attracted to distractor A, that has 1-1 as numerator, and the denominator is the smaller denominator subtracted from the larger denominator of the two fractions in the question. The incorrect student responses of students from the European and

Asian countries tended to avoid distractor A, but were attracted to either distractor B or distractor C (Figure 3). The magnitude of incorrect student responses (Table S1) and attraction to specific distractors within each country (Figure 3; Table S2) may point to Nordic students believing that whole-number arithmetic may be valid for fractions (Gabriel et al., 2013; Siegler et al., 2011). Natural number bias often lead students to make errors when dealing with fractions (Lamon, 2006; Ni & Zhou, 2005), and students tend to believe that a fraction is an arrangement of two parts, not a single quantity (Yeo & Webel, 2022). The magnitude of the Nordic students choosing distractor A over the correct answer also points to a lack of understanding the symbolic notation of fractions (i.e., do not understand the relationship between numerator and denominator), although they may have some understanding of part-whole relationships. Lacking proper conceptual understanding of fractions and the sub-construct part-whole (Kieren, 1976), may have influenced the performance of the students on the focus item and its clones. In addition, Nordic students may struggle with the procedural construct of additive operations with fractions (Haryanto, 2019).

The question then is why Nordic students show inferior performance on this type of fraction items, even across several cycles of the large-scale assessment. One reason may be cross-cultural differences in instruction (Charalambous et al., 2010; Klacznski et al., 2019). Students performance on fractions is influenced by their mathematics' teachers' content knowledge (Bobis, 2011) and how the teachers present the material (Charalambous et al., 2010). Cross-cultural differences in teacher training, educational background, requirements and alike, would most likely play a role in how and what is taught in different classrooms in different countries. Across nations, there are also differences in how textbooks used in classrooms present the topic of fractions (Charalambous et al., 2010). Learning opportunities and expectations to students as manifested in for example fraction tasks in textbooks vary across cultures (Charalambous et al., 2010), and maybe the way the TIMSS item is presented is unfamiliar to how addition and subtraction of fractions are taught in Nordic classrooms. That Nordic students show inferior performance on one specific type of fraction items, warrants a cross-cultural study of textbooks and how addition and subtraction of fractions are presented, to see if the noticed inferior performance of Nordic students on this type of fraction items may be due to the means of fraction instruction in Nordic classrooms.

In addition, student opportunities to learn is influenced by the instructional time spent on mathematics. In TIMSS 2019, an average grade eight student received 137 hours of mathematics instruction a year (13 % of the average annual instructional hours of 1,023), ranging from 200 in Chile to 102 in Cyprus (Mullis et al., 2020). The Nordic countries had less annual instructional time spent on mathematics than most countries participating in TIMSS 2019, with 105, 108, and 111 hours, for Sweden, Norway, and Finland, respectively (Mullis et al., 2020). The lower amount of time spent on instruction in mathematics the Nordic classrooms compared to those of other countries participating in TIMSS 2019, would most likely affect all types of items in mathematics, not just a specific type of item. Comparing annual instructional hours to the overall average mathematics achievement for each participating country, the Nordic countries in general do well, being ranked as 14, 15 and 16 out of 39 countries in mathematics achievement while being among the 9 countries with the lowest annual instructional time spent on mathematics (Mullis et al., 2020).

Item Difficulty as a Function of Item Features

That Nordic students' struggle with one type of fraction item across several cycles of TIMSS, triggered an investigation into whether this observation was generalizable to fraction items in general or maybe to more specific common denominator items. In Norway, teachers at secondary schools and upper secondary schools have reported that students struggle with fractions in general (Lande, 2022), and a master thesis shows that Norwegian grade 8 students showed deficit performance on four out of six selected fraction items from TIMSS 2003 (Utgård, 2008). The item showing the largest deficit performance for Norwegian students involved the concept of common denominator (Utgård, 2008), just like the item identified in the blogpost (Appendix III).

Quantifying the item variance and the person variance on TIMSS 2019 data, 15-25 % of the variance was attributed to difference in difficulty between the items and 19-36 % was attributed to the individual differences in ability between students (Table 2). For Norway, Sweden, and Italy, the item and person variances were balanced, i.e., item responses attributed equally to the item characteristics and to the student ability. In England and Lithuania, item response dependent slightly more on individual differences in ability between students than on which items the students responded to. These results are in line with the notion that the Nordic and the Anglosphere countries show larger differences in the student performance levels (Marcq & Braeken, 2022). In Korea and Singapore person variance was roughly the double the item variance. This is opposed to previous studies showing that «on average, across 77 countries, item variance was roughly the double the person variance» (Marcq & Braeken, 2022, p. 342) for PISA 2018 items, and «the variance components of the hierarchical item response model showed that about 30% of the variation in responses was due to the item characteristics, compared with only 15% due to the pupil abilities» for TIMSS 2011 (Daus et al., 2019, p.1108).

These studies are more in line with the results from Finland where item responses dependent slightly more on which items the students responded to (24.8 %) than on individual differences in ability between students (19.9 %). These differences between the countries likely point to differences in educational systems, and to differences in, for example, curriculum, learning goals, textbooks used, teacher training, and what is taught, and the findings coud serve as motivation for research into understanding regional trends in education.

This study implies that Nordic students did not show inferior performance on either fractions and probability items in general or fraction items requiring a common denominator in TIMSS 2019, compared to non-fraction items. As shown by the model comparison indices, the cross-classified mixed effect models show no significant differences between the null models and the nested fraction-and-common-denominator models within each country (Table 2), with item features only accounting for only 0.1 – 0.2 % of the item variation. The pattern is the same for all the countries included, across Nordic and non-Nordic countries, and independent of whether it is a high – or low scoring country. Thus, students within each country do not show inferior performance on fraction items, or fractions items involving common denominator, compared to other items in mathematics in TIMSS 2019. This study does not support the notion that fraction items or common denominator items are systematically more difficult than other items in mathematics in TIMSS. Despite a vast number of research articles on fractions describing students difficulties in understanding fractions and their operations (for example: Ni & Zhou, 2005; Vamvakoussi, 2015; Moss & Case, 1999; Siegler & Pyke, 2013), fractions may just be one type of items students find difficult in mathematics. Herein the items categorized as belonging to the category of fractions comprise a number of different types of tasks (see Method section for more details), in accordance with Kieren's sub-constructs (Kieren, 1976), including part-whole comparisons, quotients, ratios, operators, and measures on a number line, and involving both conceptual and procedural knowledge (Bailey et al., 2014). Therefore, different types of fraction items may show different difficulty for students within different countries, as observed on one specific type of item as mentioned in the blogpost (Appendix III).

Within countries there are different profiles of relative weaknesses and strengths, i.e. different systematic response variation across items (Marcq & Braeken, 2022). This is visualized in Landscape plots (Braeken & Paap, 2020; Navarro et al., 2004) where items are ranked according to actual difficulty within each country (Figure 4; Figures S3-S7). The items are color-coded so that items in mathematics are grey, fraction items are blue, and common denominator items are red. For example, the colors are spread out and neither fraction items nor common denominator items are clustered among the most difficult items for Norwegian students (Figure 4; upper left diagram), or students from other countries (Figures S3-S7). This agrees with the results from the cross-classified mixed effect models (Table 2) where no significant differences between item categories were found. None of the Landscape plots, the direct pairwise comparions of the item difficulty ranking differences between countries, show a systematic pattern within a country for either fraction or common denominator items, but there are some regional themes (Figure 4; Figures S3-7). The more clustered along the diagonal, the more similar the country profile of relative strengths and weaknesses on the items in TIMSS 2019 to the Norwegian profile. The most similar profiles to the Norwegian profile belong to Sweden and Finland, then England before Lithuania and Italy, while the profiles of Singapore and Korea have the most different profiles (i.e., the items are the most spread out from the diagonal). The gradient of differences seems to match the cultural similarities and differences between the herein selected countries, with the Asian countries being the most distant culturally, and having the most different systematic response variation across items to Norway. The Asian countries are also the countries with the highest overall average mathematics achievement, and it would therefore be expected that they show a different pattern from that of the average scoring Nordic countries.

Limitations

TIMSS data are cross sectional data collected at a specific moment in time, but the data are of high quality as representative samples are carefully selected within each participating country using standardized procedures (Martin et al., 2020). The items issued are designed to fit cross-country curriculums, and tested in a pilot study across the participating countries the year before the actual assessment takes place. There is no reason to believe that the data used herein will be unrepresentative of the populations they are representing. In addition, the focus has been on the item population, not the person population. However, the items used are not part of an experimental set of items, and only data from grade 8 students have been used. Further studies should validate the findings across age groups.

Herein, the assumption has been that every item has the same discrimination, thus, a 1 PL IRT model has been employed. This has been done to be able to compare the difficulty parameters among themselves as the discrimination is set to be equal. The focus herein has been on the item population and variance components of students and items; therefore, sample weights have been ignored. If the focus had been on the persons, and not the items, sample weights should have been included into the analyses.

The results show that Nordic students underperform on one specific type of fraction items, but that this is not generalizable to fraction items in general nor common denominator items. The mechanism behind the Nordic struggle with the focus item has yet to be found.

Conclusion

The present study was motivated by a desire to validate the claim that Nordic students underperform on fraction items. Focusing on the item population and variance components of students and items on TIMSS data, a probability model showing proportion of correct item answers and an explorative cross-classified mixed IRT model, did not support the claim. Nordic students show inferior performance on one type of fraction item across time, but this is not generalizable to neither common denominator items nor fraction items in general. The study contributes to an understanding of the importance of items and its role in investigating response variation sources. Investigating the within country response variation of items may be one way to uncover potential information (Marcq & Braeken, 2022), and to produce value-added information for curriculum designers, test developers, and teachers by providing more detailed performance profiles of each country (Daus et al., 2019). Different countries have different country profiles, and including item difficulty into educational research on international large-scale assessments can help us understand more of the complexities of educational systems.

References

Bailey, D. H., Siegler, R. S., & Geary, D. C. (2014). Early predictors of middle school fraction knowledge. *Developmental Science*, *17*(5), 775–785. https://doi.org/10.1111/desc.12155

Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 1–33). Lawrence Erlbaum Associates Publishers.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Behr, M. J., Harel, G., Post, T., & Lesh, R. A. (1993). Rational numbers: Toward a semantic analysis. In T. P. Carpenter, E. Fennema, & T. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 13–47). Erlbaum.

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals*. Longman Inc.

Bobis, J. (2011). Fractions: Best evidence and its implications for practice. In J. Way & J. Bobis (Eds.), *Fractions: Teaching for Understanding* (pp. 13–22). Australian Association of Mathematics Teachers.

Booth, J. L., & Newton, K. J. (2012). Fractions: Could they really be the gatekeeper's doorman? *Contemporary Educational Psychology*, *37*(4), 247–253. https://doi.org/10.1016/j.cedpsych.2012.07.001

Booth, J. L., Newton, K. J., & Twiss-Garrity, L. K. (2014). The impact of fraction magnitude knowledge on algebra performance and learning. *Journal of Experimental Child Psychology*, *118*, 110–118. https://doi.org/10.1016/j.jecp.2013.09.001

Braeken, J., & Paap, M. C. S. (2020). Making Fixed-Precision Between-Item Multidimensional Computerized Adaptive Tests Even Shorter by Reducing the Asymmetry Between Selection and Stopping Rules. *Applied Psychological Measurement*, *44*(7–8), 531–547. https://doi.org/10.1177/0146621620932666

Braithwaite, D. W., Pyke, A. A., & Siegler, R. (2017). A Computational Model of Fraction Arithmetic. *Psychological Review*, *124*(5), 603–625. https://dx.doi.org/10.1037/rev0000072

Cajori. (1894). *A History of Mathematics*. MACMILLAN ANC CO. http://archive.org/details/historyofmathema001062mbp

Charalambous, C. Y., Delaney, S., Hsu, H.-Y., & Mesa, V. (2010). A Comparative Analysis of the Addition and Subtraction of Fractions in Textbooks from Three Countries. *Mathematical Thinking and Learning*, *12*(2), 117–151. https://doi.org/10.1080/10986060903460070

Charalambous, C. Y., & Pitta-Pantazi, D. (2007). Drawing on a Theoretical Model to Study Students' Understandings of Fractions. *Educational Studies in Mathematics*, *64*(3), 293–316. https://doi.org/10.1007/s10649-006-9036-2

Chinnappan, M., & Forrester, T. (2014). Generating procedural and conceptual knowledge of fractions by pre-service teachers. *Mathematics Education Research Journal*, *26*(4), 871–896. https://doi.org/10.1007/s13394-014-0131-x

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Daus, S., Nilsen, T., & Braeken, J. (2019). Exploring Content Knowledge: Country Profile of Science Strengths and Weaknesses in TIMSS. Possible Implications for Educational Professionals and Science Research. *Scandinavian Journal of Educational Research*, *63*(7), 1102–1120. https://doi.org/10.1080/00313831.2018.1478882

De Boeck, P. (2008). Random Item IRT Models. *Psychometrika*, *73*(4), 533–559. https://doi.org/10.1007/s11336-008-9092-x

De Boeck, P., & Wilson, M. (Eds.). (2011). *Explanatory item response models: A generalized linear and nonlinear approach* (Softcover reprint of the hardcover 1. ed. 2004). Springer.

DeWolf, M., Bassok, M., & Holyoak, K. J. (2015). From rational numbers to algebra: Separable contributions of decimal magnitude and relational understanding of fractions. *Journal of Experimental Child Psychology*, *133*, 72–84. https://doi.org/10.1016/j.jecp.2015.01.013

DeWolf, M., & Vosniadou, S. (2015). The representation of fraction magnitudes and the whole number bias reconsidered. *Learning and Instruction*, *37*, 39–49. https://doi.org/10.1016/j.learninstruc.2014.07.002

Durkin, K., & Rittle-Johnson, B. (2015). Diagnosing misconceptions: Revealing changing decimal fraction knowledge. *Learning and Instruction*, *37*, 21–29. https://doi.org/10.1016/j.learninstruc.2014.08.003

Empson, S. B., Levi, L., & Carpenter, T. P. (2011). The Algebraic Nature of Fractions: Developing Relational Thinking in Elementary School. In J. Cai & E. Knuth (Eds.), *Early Algebraization: A Global Dialogue from Multiple Perspectives* (pp. 409–428). Springer. https://doi.org/10.1007/978-3-642-17735-4_22

Faulkenberry, T. J. (2013). The conceptual/procedural distinction belongs to strategies, not tasks: A comment on Gabriel et al. (2013). *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00820

Fish, D. W. (1874). *The Complete Arithmetic*. Ivison, Blakeman, Taylor & Company.

Fishbein, B., Foy, P., & Yin, L. (2021). TIMSS 2019 user guide for the international database. *Https://Timssandpirls. Bc. Edu/Timss2019/International-Database*.

Foy, P., Arora, A., & Stanco, G. M. (Eds.). (2013). *Timss 2011 user guide for the international database*. TIMSS & PIRLS International Study Center.

Foy, P., & Olson, J. F. (Eds.). (2009). *TIMSS 2007 International Database and User Guide*. TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2007/idb_ug.html

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*, 715. https://doi.org/10.3389/fpsyg.2013.00715

Hackenberg, A. J. (2013). The fractional knowledge and algebraic reasoning of students with the first multiplicative concept. *The Journal of Mathematical Behavior*, *32*(3), 538–563. https://doi.org/10.1016/j.jmathb.2013.06.007

Hanson, S. A., & Hogan, T. P. (2000). Computational Estimation Skill of College Students. *Journal for Research in Mathematics Education*, *31*(4), 483–499. https://doi.org/10.2307/749654

Haryanto, B. (2019). *Assessing Learning Progression in the Domain of Fractions* [Doctoral dissertation]. Flinders University, College of Education, Psychology and Social Work.

Hurst, M. A., & Cordes, S. (2018). Children's understanding of fraction and decimal symbols and the notation-specific relation to pre-algebra ability. *Journal of Experimental Child Psychology*, *168*, 32–48. https://doi.org/10.1016/j.jecp.2017.12.003

Kaarstein, H., Radišić, J., Lehre, A.-C., Nilsen, T., & Bergen, O. K. (2020). *TIMSS 2019. Kortrapport*. Universitetet i Oslo.

Kieren, T. E. (1976). On the matehematical, cognitive, and instructional foundations of rational numbers. In R. A. Lesh & D. A. Bradbard (Eds.), *Number and Measurement: Papers from a Research Workshop* (pp. 101–144).

Kieren, T. E. (1993). Rational and fractional numbers: From quotient fields to recursive understanding. In *Rational numbers: An integration of research* (pp. 49–84). Lawrence Erlbaum Associates, Inc.

Klacznski, P. A., Amsel, E. A., & Felmban, W. S. (2019). Age, numeracy, and cultural differences in Chinese and American adolescents' performance on the ratio bias task. *Journal of Experimental Child Psychology*, *188*(Article 104669). https://doi.org/10.1016/j.jecp.2019.104669

Lamon, S. J. (2006). *Teaching Fractions and Ratios for Understanding: Essential Content Knowledge and Instructional Strategies for Teachers*. Routledge.

Lande, D. (2022). *Illustrasjon av brøk: En kvantitativ undersøkelse blant lærer på 4.-7. Trinn* [Nord Universitetet]. https://nordopen.nord.no/nord-xmlui/bitstream/handle/11250/3048263/LandeDavid.pdf?sequence=1&isAllowed=y

LaRoche, S., Joncas, M., & Foy, P. (2020). Sample design in TIMSS 2019. In *Methods and Procedures: TIMSS 2019 Technical Report* (p. 3.1-3.33). Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/methods/chapter-3.html

Lortie-Forgues, H., Tian, J., & Siegler, R. S. (2015). Why is learning fraction and decimal arithmetic so difficult? *Developmental Review*, *38*, 201–221. https://doi.org/10.1016/j.dr.2015.07.008

Marcq, K., & Braeken, J. (2022). The blind side: Exploring item variance in PISA 2018 cognitive domains. *Assessment in Education: Principles, Policy & Practice*, *29*(3), 332–360. https://doi.org/0.1080/0969594X.2022.2097199

Martin, M. O. (Ed.). (2005). *TIMSS 2003 User Guide for the International Database*. TIMSS & PIRLS International Study Center, Boston College. https://timss.bc.edu/timss2003i/userguide.html

Martin, M. O., Davier, M. von, & Mullis, I. V. S. (Eds.). (2020). *Methods and Procedures: TIMSS2019 Technical Report*. TIMSS & PIRLS International Study Center.

Merzbach, U., & Boyer, C. (2011). *A History of Mathematics* (3rd ed.). John Wiley & Sons. Inc.

Moss, J., & Case, R. (1999). Developing Children's Understanding of the Rational Numbers: A New Model and an Experimental Curriculum. *Journal for Research in Mathematics Education*, *30*(2), 122–147. https://doi.org/10.2307/749607

Mostert, T. M. M., & Hickendorff, M. (2023). Pizzas or no pizzas: An advantage of word problems in fraction arithmetic? *Learning and Instruction*, *86*, 101775. https://doi.org/10.1016/j.learninstruc.2023.101775

Mou, Y., Li, Y., Hoard, M. K., Nugent, L. D., Chu, F. W., Rouder, J. N., & Geary, D. C. (2016). Developmental Foundations of Children's Fraction Magnitude Knowledge. *Cognitive Development*, *39*, 141–153. https://doi.org/10.1016/j.cogdev.2016.05.002

Mullis, I. V. S., & Martin, M. O. (Eds.). (2013). *Timss 2015 Assessment Frameworks*. TIMSS & PIRLS International Study Center, Lynch School ofEducation, Boston College and International Association for the Evaluation of Educational Achievement.

Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. TIMSS & PIRLS International Study Center ; International Association for the Evaluation of Educational Achievement. https://files.eric.ed.gov/fulltext/ED596167.pdf

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (Eds.). (2012). *TIMSS 2011 International Results in Mathematics*. TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2011/downloads/T11_IR_M_Chapter1.pdf

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. TIMSS & PIRLS International Study Center at Boston College website. https://timss2019.org/reports

National Mathematics Advisory Panel. (2008). *Foundations for success: Final report of the national mathematics advisory panel.* US Department of Education.

Navarro, D. J., Pitt, M. A., & Myung, I. J. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology*, *49*(1), 47–84. https://doi.org/10.1016/j.cogpsych.2003.11.001

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

Park, J., Güçler, B., & McCrory, R. (2013). Teaching prospective teachers about fractions: Historical and pedagogical perspectives. *Educational Studies in Mathematics*, *82*(3), 455–479. https://doi.org/10.1007/s10649-012-9440-8

Philpot, R., Lindquist, M., Mullis, I. V. S., & Aldrich, C. E. A. (2023). *TIMSS 2023 Mathematics Framework*.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Rasch, G. (Georg). (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago : University of Chicago Press. http://archive.org/details/probabilisticmod0000rasc

Raths, L. E. (1932). The Grade-Placement of Addition and Subtraction of Fractions. *Educational Research Bulletin*, *11*(2), 29–38.

Rocher, T., & Hastedt, D. (2020). *International large-scale assessments in education: A brief guide* (10; IEA Compass: Briefs in Education). IEA. https://www.iea.nl/sites/default/files/2021-07/2021.07.20_ILSAs%20in%20education%20-%20a%20brief%20guide%20Compass%2010.pdf

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Susperreguy, M. I., & Chen, M. (2012). Early Predictors of High School Mathematics Achievement. *Psychological Science*, *23*(7), Article 7. https://doi.org/10.1177/0956797612440101

Siegler, R. S., & Lortie-Forgues, H. (2015). Conceptual knowledge of fraction arithmetic. *Journal of Educational Psychology*, *107*(3), 909–918. https://doi.org/10.1037/edu0000025

Siegler, R. S., & Pyke, A. A. (2013). Developmental and individual differences in understanding of fractions. *Developmental Psychology*, *49*(10), 1994–2004. https://doi.org/10.1037/a0031200

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology*, *62*(4), 273–296. https://doi.org/10.1016/j.cogpsych.2011.03.001

Stafylidou, S., & Vosniadou, S. (2004). The development of students' understanding of the numerical value of fractions. *Learning and Instruction*, *14*(5), 503–518. https://doi.org/10.1016/j.learninstruc.2004.06.015

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

Tukey, J. W. (1977). *Exploratory Data Analysis*. Pearson.

Tunc Pekkan, Z. (2008). *Modeling grade eight students construction of fraction multiplying schemes and algebraic operations* [Doctoral dissertation, The University of Georgia]. https://getd.libs.uga.edu/pdfs

Tyler, R. W. (1949). *Basic principles of curriculum and instruction.* The University of Chicago Press, Ltd.

Utgård, P.-A. (2008). *Elevers forståelse av brøkbegrepet. En sekundæranalyse av TIMSS-undersøkelsen i 2003 på 8. Klassetrinn* [Master thesis, University of Oslo]. http://urn.nb.no/URN:NBN:no-19748

Vamvakoussi, X. (2015). The development of rational number knowledge: Old topic, new insights. *Learning and Instruction*, *37*, 50–55. https://doi.org/10.1016/j.learninstruc.2015.01.002

Van Dooren, W., Lehtinen, E., & Verschaffel, L. (2015). Unraveling the gap between natural and rational numbers. *Learning and Instruction*, *37*, 1–4. https://doi.org/10.1016/j.learninstruc.2015.01.001

Vocabulary.com. (n.d.). Vocabulary.Com. Retrieved November 12, 2023, from https://www.vocabulary.com/

von Davier, M. (2016). The Rasch model. In W. J. van der Linden (Ed.), *Handbook of item response theory* (2nd ed., Vol. 1, pp. 31–48). CRC Press. https://www.routledgehandbooks.com/doi/10.1201/9781315374512-5

Yeo, S., & Webel, C. (2022). Elementary students' fraction reasoning: A measurement approach to fractions in a dynamic environment. *Mathematical Thinking and Learning*, 1–27. https://doi.org/10.1080/10986065.2022.2025639

Appendix I

NOTIFICATION FORM GDPR – Norwegian Centre for Research data (NSD)

NSD

Result of Notification Test: Not Subject to Notification

You have indicated that neither directly or indirectly identifiable personal data will be registered in the project.

If no personal data is to be registered, the project will not be subject to notification, and you will not have to submit a notification form.

Please note that this is a guidance based on information that you have given in the notification test and not a formal confirmation.

For your information: *In order for a project not to be subject to notification, we presuppose that all information processed using electronic equipment in the project remains anonymous.*

Anonymous information is defined as information that cannot identify individuals in the data set in any of the following ways:

- *directly, through uniquely identifiable characteristic (such as name, social security number, email address, etc.)*
- *indirectly, through a combination of background variables (such as residence/institution, gender, age, etc.)*
- *through a list of names referring to an encryption formula or code, or*
- *through recognizable faces on photographs or video recordings.*

Furthermore, we presuppose that names/consent forms are not linked to sensitive personal data.

Kind regards,
NSD Data Protection

NSD – Norsk senter for forskningsdata AS Harald Hårfagres gate 29 Tel: +47-55 58 21 17 nsd@nsd.no Org.nr. 985 321 884
NSD – Norwegian Centre for Research Data NO-5007 Bergen, NORWAY Faks: +47-55 58 96 50 www.nsd.no

Appendix II

Data management and analysis code

Software used: R version 4.3.1. (R Core Team, 2021)

Data

Data are publicly available and retrievable at https://timssandpirls.bc.edu/databases-landing.html.

R syntax for data management steps and analyses

```
##########################################################################################
###########                    FRACTION PROJECT                    ###########
##########################################################################################

setwd("~")

library("foreign")
ISO <- c(
  "NOR",
  "SWE",
  "FIN",
  "ENG",
  "ITA",
  "LTU",
  "SGP",
  "KOR"
)

Beta <- vector(mode = "list", length = length(ISO))
B0 <- vector(mode = "list", length = length(ISO))
B1 <- vector(mode = "list", length = length(ISO))
V0 <- vector(mode = "list", length = length(ISO))
V1 <- vector(mode = "list", length = length(ISO))
R2 <- vector(mode = "list", length = length(ISO))
ModelComparison <- vector(mode = "list", length = length(ISO))
NPI <- vector(mode = "list", length = length(ISO))
```

```
for (country in 1:length(ISO)) {
  ###########          LOOP THROUGH COUNTRIES USING ISO acronym TIMSS database          ###########

        ###########################################################################################
             ###########                    DATA PREPARATION                    ###########
        ###########################################################################################
  datafile <- tolower(paste0("./data/bsa", ISO[country], "m7.sav"))
  data <- read.spss(datafile, to.data.frame = TRUE, use.missings = FALSE, use.value.labels = FALSE)
  dim(data)
  LAB.bg <- c(
                          "IDCNTRY", "IDBOOK", "IDSCHOOL", "IDCLASS", "IDSTUD",
                          "IDPOP", "ITSEX", "TOTWGT", "VERSION", "SCOPE"
        )
  # Only Math items = start with M
  # Norway only has electronic ME: second character is then an E
        #(P for paper, N for easy-TIMSS, Q for problem-solving)
  # Don't include the freq of screen visits & time spent
  LAB.items <- grep("_(F|S)$", grep("^ME", names(data), value = TRUE), invert = TRUE, value = TRUE)
  data <- data[, c(LAB.bg, LAB.items)]
  dim(data)
  # NOT administered = systematically missing so fine
  # Omitted = 9 or 99
  data[data == 9 | data == 99] <- NA
  # SCORING first MC: 4 options then CR
  # Based on TIMSS supplied SAS program
  # A = 1 => correct
  A <- c(
    "ME52092", "ME52418B", "ME72007B", "ME72007C", "ME72007E", "ME72076", "ME72180A", "ME72170B",
              "ME62164", "ME62084", "ME72178B", "ME72178C", "ME72178E", "ME72020C", "ME72027", "ME72164A",
              "ME72164D", "ME52413","ME52407", "ME62219", "ME62342", "ME52068", "ME72055B", "ME72055D",
              "ME72055F", "ME72090", "ME62245", "ME62345BB", "ME62345BD", "ME72125", "ME72232A", "ME72232D",
              "ME62095", "ME62076", "ME62146", "ME62242", "ME62048A", "ME72080", "ME72081A", "ME72081C",
              "ME72140B", "ME72140D", "ME72140F", "ME72192"
  )
  table(data[, A[1]])
  data[, A][data[, A] > 1 & data[, A] <= 4] <- 0
  data[, A][data[, A] == 1] <- 1
  table(data[, A[1]])
```

```
# B = 2 => correct
B <- c(
  "ME52024", "ME52072", "ME52161", "ME72007A", "ME72007D", "ME72025", "ME72180B", "ME72170C",
            "ME62351", "ME62174", "ME62132B", "ME72178A", "ME72178D", "ME72020D", "ME72067", "ME72164B",
            "ME72164C", "ME72164E", "ME52078", "ME52130", "ME62335", "ME62123B", "ME52079", "ME52147",
            "ME52067", "ME72055A", "ME72055C", "ME72055E", "ME72222", "ME72233", "ME62329", "ME62212",
            "ME62284", "ME62345BA", "ME62345BC", "ME72232B", "ME72232C", "ME62230", "ME72189", "ME72221",
            "ME72211A", "ME62001", "ME62048B", "ME62048C", "ME72223", "ME72081B", "ME72081D", "ME72140A",
            "ME72140C", "ME72140E"
)
table(data[, B[1]])
data[, B][data[, B] == 1 | (data[, B] <= 4 & data[, B] >= 3)] <- 0
data[, B][data[, B] == 2] <- 1
table(data[, B[1]])
# C = 3 => correct
C <- c(
  "ME52063", "ME52083", "ME52082", "ME72068", "ME72180C", "ME72170A", "ME62005", "ME62261",
            "ME72020A", "ME72020B", "ME52034", "ME52073", "ME62040", "ME62123A", "ME52204", "ME52419A",
            "ME52419B", "ME72188", "ME62115", "ME72022", "ME72013", "ME62194", "ME72043", "ME72150",
            "ME62067", "ME62120", "ME72005", "ME72154"
)
table(data[, C[1]])
data[, C][data[, C] == 4 | (data[, C] < 3 & data[, C] >= 1)] <- 0
data[, C][data[, C] == 3] <- 1
table(data[, C[1]])
# D = 4 => correct
D <- c(
  "ME52125", "ME52046", "ME52418A", "ME72103", "ME62223", "ME72234", "ME72083B", "ME52134",
            "ME52502D", "ME52426", "ME62149", "ME62133", "ME52208", "ME52115", "ME72172", "ME62350",
            "ME72038", "ME72237", "ME62271", "ME62171", "ME72211B", "ME62320", "ME72220", "ME62341"
)
table(data[, D[1]])
data[, D][data[, D] < 4 & data[, D] >= 1] <- 0
data[, D][data[, D] == 4] <- 1
table(data[, D[1]])
# CR
CR <- c(
  "ME52058A", "ME52058B", "ME52229", "ME52146A", "ME52146B", "ME72007", "ME72017", "ME72190",
```

```
                    "ME72056", "ME72098", "ME72121", "ME72180", "ME72198", "ME72198A", "ME72198B", "ME72227",
                    "ME72170", "ME72209", "ME62139", "ME62142", "ME62027", "ME62244", "ME62244A", "ME62244B",
                    "ME62300", "ME62254", "ME62132A", "ME72178","ME72020", "ME72052", "ME72052A", "ME72052B",
                    "ME72083A", "ME72108A", "ME72108B", "ME72181", "ME72126", "ME72164", "ME72185A", "ME72185B",
                    "ME52174A", "ME52174B", "ME52110", "ME52105", "ME52036", "ME52502", "ME52502A", "ME52502B",
                    "ME52502C", "ME52117", "ME62150", "ME62002", "ME62241", "ME62105", "ME62288", "ME62288A",
  "ME62288B", "ME62173", "ME52364", "ME52215", "ME52087", "ME52087A", "ME52087B", "ME52048",
                    "ME52039", "ME52421", "ME72002", "ME72035", "ME72055", "ME72106A","ME72106B", "ME72106C",
                    "ME72128A", "ME72128B", "ME72119", "ME72153A", "ME72153B", "ME62151", "ME62346", "ME62056",
                    "ME62317", "ME62317A", "ME62317B", "ME62317C", "ME62078", "ME62287", "ME62345A", "ME62345AA",
                    "ME62345AB", "ME62345AC", "ME62345AD", "ME62345B", "ME72187", "ME72045", "ME72049", "ME72069",
                    "ME72074", "ME72095", "ME72095A", "ME72095B", "ME72109", "ME72196", "ME72232", "ME72206",
                    "ME62152", "ME62215", "ME62215A", "ME62215B", "ME62143", "ME62030", "ME62301", "ME62344",
  "ME62296", "ME72001", "ME72019", "ME72024", "ME72225", "ME72225A", "ME72225B", "ME72110A",
                    "ME72110B", "ME72139", "ME72229", "ME72171", "ME62214", "ME62154", "ME62250A", "ME62250B",
                    "ME62170", "ME62170A", "ME62170B", "ME62192", "ME62072", "ME62048", "ME72021", "ME72026",
                    "ME72041A", "ME72041B", "ME72094", "ME72059", "ME72081", "ME72140", "ME72120", "ME72131",
                    "ME72147", "ME72161"
)
table(data[, CR[1]])
          # no credit
data[, CR][data[, CR] >= 70 & data[, CR] < 80] <- 0
          # partial credit recoded as 1 for binary score as unclear what "partial" means and can be "full at times"
data[, CR][data[, CR] >= 10 & data[, CR] < 20] <- 1
          # full credit recoded as 1 for binary score
data[, CR][data[, CR] >= 20 & data[, CR] < 30] <- 1
table(data[, CR[1]])
# Not Reached = 6 or 96          --> for item calibration better to treat as NA (see e.g., Mislevy),
          # for person scoring as incorrect
data[, LAB.items][data[, LAB.items] == 6 | data[, LAB.items] == 96] <- NA

                                        ##### ITEMS TO KEEP/KICK OUT#####
# Some derived items already incorporated in the dataset, the parts kicked out,
          # except those that don't have any parent item
LAB.items <- grep("\\d$", LAB.items, value = TRUE)
LAB.items <- c(
  LAB.items, "ME52058A", "ME52058B", "ME52146A", "ME52146B", "ME52418A", "ME52418B", "ME62132A",
```

```
                    "ME62132B", "ME72083A", "ME72083B", "ME72108A", "ME72108B", "ME72185A", "ME72185B",
"ME52174A",
                    "ME52174B", "ME52419A", "ME52419B", "ME72106A", "ME72106B", "ME72106C", "ME72128A",
"ME72128B",
                    "ME72153A", "ME72153B", "ME62345A", "ME62345B", "ME72110A", "ME72110B", "ME72211A",
"ME72211B",
                    "ME62250A", "ME62250B", "ME72041A", "ME72041B"
  )
  # number of items for a country
          length(LAB.items)
  data <- data[, c(LAB.bg, LAB.items)] # clean dataset

  # Quick checks based on descriptives official stats
  # summary(data[,LAB.bg])
  # Everyone has a studentweight given comments in other code
  # Number of unique student id's vs number of rows
  # c(length(unique(data$IDSTUD)),nrow(data))
  # Number of observed item responses per student
  # boxplot(apply(!is.na(data[,LAB.items]),1,sum))
  # Those students with less than 5 responses
  #       data[apply(!is.na(data[,LAB.items]),1,sum)<5,LAB.bg]
  # Throw out students with less than 5 responses
  data <- data[apply(!is.na(data[, LAB.items]), 1, sum) >= 5, ]
  ########## CAN USE THIS DATASET FOR DESCRIPTIVES OR OTHER PURPOSES
  saveRDS(data, file = paste0("TIMSS2019", ISO[country], ".rds"))


              ##########################################################################################
               ###########     RASCH/1PL BASED ANALYSIS (note easiness parameters here)          ###########
              ##########################################################################################
  library("lme4")

  ITEMS <- data.frame(readxl::read_excel("MAE_TIMSS_G9_Items_kat_ACL_v4.xlsx", sheet = "TIMSS_2019_9M"))
  str(ITEMS)
  # Cleanup this file: Complete coding scheme and shorten variable names
  names(ITEMS)[c(1, 12:13)] <- c("Item", "CoDenom", "Spec") # shorten
  ITEMS <- ITEMS[, 1:14] # keep entry columns
  ITEMS[, 12:13][is.na(ITEMS[, 12:13])] <- 0 # complete
  ITEMS$CAT <- 1 + ITEMS$Fraction + ITEMS$CoDenom + ITEMS$Spec
```

```
# WIDE2LONG
DATA <- reshape(
                        data, varying = list(LAB.items), v.names = "Y", idvar = "IDSTUD",
                        timevar = "Item", times = LAB.items, direction = "long"
            )
# DATA[DATA$IDSTUD==50010106,c("Item","Y")]
# data[data$IDSTUD==50010106,DATA[DATA$IDSTUD==50010106,c("Item")]]
# save some working memory
            rm(data)
DATA <- DATA[!is.na(DATA$Y), ]
# number of observed item responses
            dim(DATA)
# number of students
            length(unique(DATA$IDSTUD))
            # number of observed item responses per student
summary(with(DATA, tapply(Y, IDSTUD, length)))
# number of observed item responses per item
            summary(with(DATA, tapply(Y, Item, length)))

# ADD ITEM FEATURES
DATA <- merge(DATA, ITEMS, by = "Item", all.x = TRUE)

# IRT
var.decomposition <- function(m) {
   VC0 <- c(unlist(lapply(VarCorr(m), diag)), (pi^2) / 3)
   ICC <- VC0 / sum(VC0)
   decomposition <- c(variance = VC0, icc = ICC)
   return(decomposition)
}
pseudoR2 <- function(m) {
   TOT <- predict(m, type = "link") # all
   RE <- predict(m, type = "link", random.only = TRUE) # only RE
   # predict(m,type="link",re.form=NA) is equivalent
                        FI <- model.matrix(m) %*% fixef(m)
   VAR.fix <- mean(scale(FI, scale = FALSE)^2)
   VAR.RE <- mean(scale(RE, scale = FALSE)^2)
   # In theory FI & RE uncorrelated, not in practice (sampling/estimation error)
```

```
VAR.TOT <- mean(scale(TOT, scale = FALSE)^2)
# Decomposition therefore fails: var(RE)+var(fix) != var(TOT)
VAR.res <- (pi^2) / 3
VAR.tot <- as.vector(VAR.fix + VAR.RE + VAR.res)
RVC <- c(fixed = VAR.fix, random = VAR.RE, residual = VAR.res, totalR2 = (VAR.fix + VAR.RE)) / VAR.tot
return(RVC)
}


#Null model
m0 <- glmer(Y ~ 1 + (1 | IDSTUD) + (1 | Item), data = DATA, family = binomial("logit"))
# summary(m0)
#Model including fractions and common denominator
m1 <- glmer(Y ~ 1 + (1 | IDSTUD) + (1 | Item) + Fraction + CoDenom, data = DATA, family = binomial("logit"))
# summary(m1)

###########################################################################################
##########                    GATHER OUTPUT                    ###########
###########################################################################################
B0[[country]] <- summary(m0)$coefficients
V0[[country]] <- var.decomposition(m0)
B1[[country]] <- summary(m1)$coefficients
V1[[country]] <- var.decomposition(m1)
R2[[country]] <- pseudoR2(m1)
ModelComparison[[country]] <- anova(m0, m1)
NPI[[country]] <- c(n.P = length(unique(DATA$IDSTUD)), n.I = length(unique(DATA$Item)), n = nrow(DATA))

Beta[[country]] = as.data.frame(ranef(m0,whichel="Item"))[,c("grp","condval","condsd")]
names(Beta[[country]])[2:3]=paste0(names(Beta[[country]])[2:3],".",ISO[country])

}

names(B1) <- ISO
lapply(B1, round, digit = 3)
lapply(V1, round, digit = 3)
```

```
##########################################################################################
##########                          ITEM PLOTS                          ###########
##########################################################################################

names(Beta)=ISO

Beta=Reduce(merge,Beta)

RANKS= Beta[,c("grp",paste0("condval.",ISO))]

RANKS[,paste0("condval.",ISO)]=sapply(Beta[,paste0("condval.",ISO)],rank)

RANKS=merge(RANKS,ITEMS,by.x="grp",by.y="Item")

COL=c("black","blue","red")[RANKS$CAT]


# size of plots

pdf("item_ranking.pdf", width = 11.7, height = 16.5, pointsize = 12)

par(mfrow = c(4, 2))

par(mar = c(5,4.5,1,1.5))


# Plot: Norway versus Norway (control)

plot(RANKS$condval.NOR,RANKS$condval.NOR,type="n",xlab="Norway",ylab="Norway",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(R
ANKS)))

abline(a=0,b=1)

text(RANKS$condval.NOR,RANKS$condval.NOR,labels=RANKS$grp,col=COL,cex=0.5)

# Plot: Norway versus Sweden (items ranked according to Norwegian "easiness")

plot(RANKS$condval.NOR,RANKS$condval.SWE,type="n",xlab="Norway",ylab="Sweden",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(R
ANKS)))

abline(a=0,b=1)

text(RANKS$condval.NOR,RANKS$condval.SWE,labels=RANKS$grp,col=COL,cex=0.5)

# Plot: Norway versus Finland (items ranked according to Norwegian "easiness")

plot(RANKS$condval.NOR,RANKS$condval.FIN,type="n",xlab="Norway",ylab="Finland",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(RA
NKS)))

abline(a=0,b=1)

text(RANKS$condval.NOR,RANKS$condval.FIN,labels=RANKS$grp,col=COL,cex=0.5)

# Plot: Norway versus England (items ranked according to Norwegian "easiness")

plot(RANKS$condval.NOR,RANKS$condval.ENG,type="n",xlab="Norway",ylab="England",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(R
ANKS)))

abline(a=0,b=1)

text(RANKS$condval.NOR,RANKS$condval.ENG,labels=RANKS$grp,col=COL,cex=0.5)

# Plot: Norway versus Lithuania (items ranked according to Norwegian "easiness")

plot(RANKS$condval.NOR,RANKS$condval.LTU,type="n",
xlab="Norway",ylab="Lithuania",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(RANKS)))

abline(a=0,b=1)
```

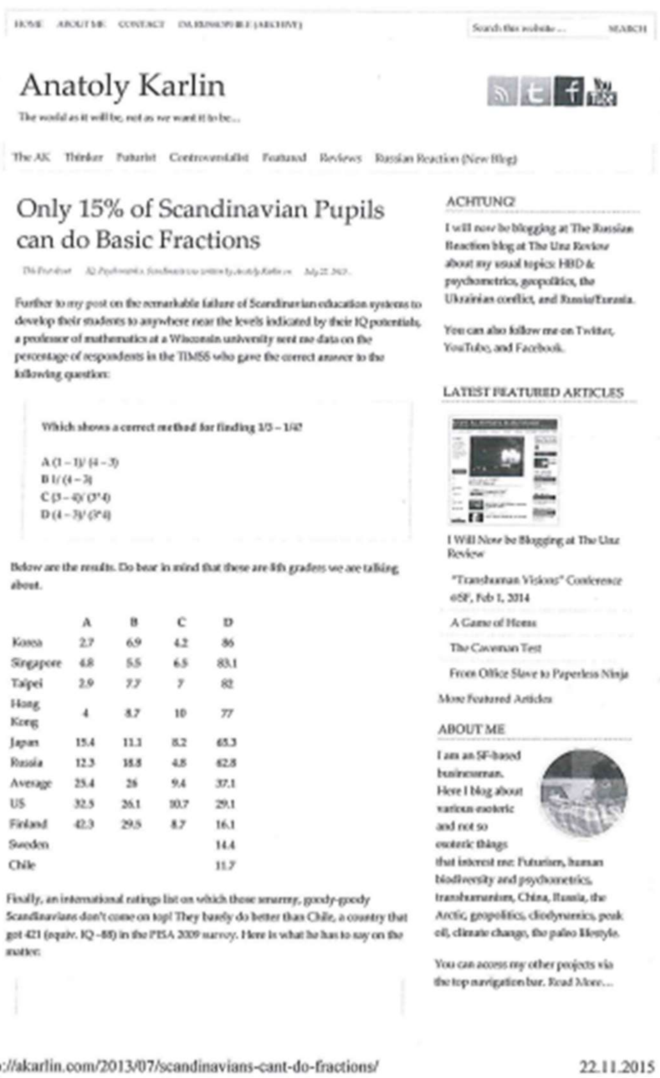
```
text(RANKS$condval.NOR,RANKS$condval.LTU,labels=RANKS$grp,col=COL,cex=0.5)

# Plot: Norway versus Italy (items ranked according to Norwegian "easiness")

plot(RANKS$condval.NOR,RANKS$condval.ITA,type="n",xlab="Norway",ylab="Italy",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(RANKS)))

abline(a=0,b=1)

text(RANKS$condval.NOR,RANKS$condval.ITA,labels=RANKS$grp,col=COL,cex=0.5)

# Plot: Norway versus Singapore (items ranked according to Norwegian "easiness")

plot(RANKS$condval.NOR,RANKS$condval.SGP,type="n",xlab="Norway",ylab="Singapore",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(RANKS)))

abline(a=0,b=1)

text(RANKS$condval.NOR,RANKS$condval.SGP,labels=RANKS$grp,col=COL,cex=0.5)

# Plot: Norway versus Korea (items ranked according to Norwegian "easiness")

plot(RANKS$condval.NOR,RANKS$condval.KOR,type="n",xlab="Norway",ylab="Korea",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(RANKS)))

abline(a=0,b=1)

text(RANKS$condval.NOR,RANKS$condval.KOR,labels=RANKS$grp,col=COL,cex=0.5)

dev.off()


### Same plots as above with 2 diagonals (+/- 50) around the correspondence line

# Plot: Norway versus Sweden (items ranked according to Norwegian "easiness") with 2 diagonals around the correspondence line

plot(RANKS$condval.NOR,RANKS$condval.SWE,type="n",xlab="Norway",ylab="Sweden",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(RANKS)))

abline(a=0,b=1)

abline(a=50,b=1,lty=2,col="grey")

abline(a=-50,b=1,lty=2,col="grey")

text(RANKS$condval.NOR,RANKS$condval.SWE,labels=RANKS$grp,col=COL,cex=0.5)

# Plot: Norway versus Finland (items ranked according to Norwegian "easiness") with 2 diagonals around the correspondence line

plot(RANKS$condval.NOR,RANKS$condval.FIN,type="n",xlab="Norway",ylab="Finland",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(RANKS)))

abline(a=0,b=1)

abline(a=50,b=1,lty=2,col="grey")

abline(a=-50,b=1,lty=2,col="grey")

text(RANKS$condval.NOR,RANKS$condval.FIN,labels=RANKS$grp,col=COL,cex=0.5)

# Plot: Norway versus England (items ranked according to Norwegian "easiness") with 2 diagonals around the correspondence line

plot(RANKS$condval.NOR,RANKS$condval.ENG,type="n",xlab="Norway",ylab="England",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(RANKS)))

abline(a=0,b=1)

abline(a=50,b=1,lty=2,col="grey")

abline(a=-50,b=1,lty=2,col="grey")

text(RANKS$condval.NOR,RANKS$condval.ENG,labels=RANKS$grp,col=COL,cex=0.5)
```

```
# Plot: Norway versus Lithuania (items ranked according to Norwegian "easiness") with 2 diagonals around the correspondence line

plot(RANKS$condval.NOR,RANKS$condval.LTU,type="n",
xlab="Norway",ylab="Lithuania",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(RANKS)))

abline(a=0,b=1)

abline(a=50,b=1,lty=2,col="grey")

abline(a=-50,b=1,lty=2,col="grey")

text(RANKS$condval.NOR,RANKS$condval.LTU,labels=RANKS$grp,col=COL,cex=0.5)

# Plot: Norway versus Italy (items ranked according to Norwegian "easiness") with 2 diagonals around the correspondence line

plot(RANKS$condval.NOR,RANKS$condval.ITA,type="n",xlab="Norway",ylab="Italy",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(RANK
S)))

abline(a=0,b=1)

abline(a=50,b=1,lty=2,col="grey")

abline(a=-50,b=1,lty=2,col="grey")

text(RANKS$condval.NOR,RANKS$condval.ITA,labels=RANKS$grp,col=COL,cex=0.5)

# Plot: Norway versus Singapore (items ranked according to Norwegian "easiness") with 2 diagonals around the correspondence line

plot(RANKS$condval.NOR,RANKS$condval.SGP,type="n",xlab="Norway",ylab="Singapore",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(R
ANKS)))

abline(a=0,b=1)

abline(a=50,b=1,lty=2,col="grey")

abline(a=-50,b=1,lty=2,col="grey")

text(RANKS$condval.NOR,RANKS$condval.SGP,labels=RANKS$grp,col=COL,cex=0.5)

# Plot: Norway versus Korea (items ranked according to Norwegian "easiness") with 2 diagonals around the correspondence line

plot(RANKS$condval.NOR,RANKS$condval.KOR,type="n",xlab="Norway",ylab="Korea",xlim=c(1,nrow(RANKS)),ylim=c(1,nrow(RA
NKS)))

abline(a=0,b=1)

abline(a=50,b=1,lty=2,col="grey")

abline(a=-50,b=1,lty=2,col="grey")

text(RANKS$condval.NOR,RANKS$condval.KOR,labels=RANKS$grp,col=COL,cex=0.5)
```

Appendix III

Supplemental material

Blog



HOME ABOUT ME CONTACT DA RUSSOPHILE (ARCHIVE) Search this website ... SEARCH

Anatoly Karlin

The world as it will be, not as we want it to be...

The AK Thinker Futurist Controversialist Featured Reviews Russian Reaction (New Blog)

Only 15% of Scandinavian Pupils can do Basic Fractions

Further to my post on the remarkable failure of Scandinavian education systems to develop their students to anywhere near the levels indicated by their IQ potentials, a professor of mathematics at a Wisconsin university sent me data on the percentage of respondents in the TIMSS who gave the correct answer to the following question:

> Which shows a correct method for finding 1/3 – 1/4?
>
> A (1 – 1)/ (4 – 3)
> B 1/ (4 – 3)
> C (3 – 4)/ (3*4)
> D (4 – 3)/ (3*4)

Below are the results. Do bear in mind that these are 4th graders we are talking about.

	A	B	C	D
Korea	2.7	6.9	4.2	86
Singapore	4.8	5.5	6.5	83.1
Taipei	2.9	7.7	7	82
Hong Kong	4	8.7	10	77
Japan	15.4	11.1	8.2	65.3
Russia	12.3	18.8	4.8	62.8
Average	25.4	26	9.4	37.1
US	32.5	26.1	10.7	29.1
Finland	42.3	29.5	8.7	16.1
Sweden				14.4
Chile				11.7

Finally, an international ratings list on which those smarmy, goody-goody Scandinavians don't come on top! They barely do better than Chile, a country that got 421 (equiv. IQ ~88) in the PISA 2009 survey. Here is what he has to say on the matter:

ACHTUNG!

I will now be blogging at The Russian Reaction blog at The Unz Review about my usual topics: HBD & psychometrics, geopolitics, the Ukrainian conflict, and Russia/Eurasia.

You can also follow me on Twitter, YouTube, and Facebook.

LATEST FEATURED ARTICLES

I Will Now be Blogging at The Unz Review

"Transhuman Visions" Conference @SF, Feb 1, 2014

A Game of Homs

The Caveman Test

From Office Slave to Paperless Ninja

More Featured Articles

ABOUT ME

I am an SF-based businessman. Here I blog about various esoteric and not so esoteric things that interest me: Futurism, human biodiversity and psychometrics, transhumanism, China, Russia, the Arctic, geopolitics, cliodynamics, peak oil, climate change, the paleo lifestyle.

You can access my other projects via the top navigation bar. Read More...

http://akarlin.com/2013/07/scandinavians-cant-do-fractions/ 22.11.2015

> One interesting fact is that among the 42 countries which tested 4th grade students, Finland had the highest percent of students who picked answer A and the third lowest percent correct. Chile had 11.7 correct and Sweden had 14.4 percent correct. The Finnish result is likely a surprise to the people who have praised the Finnish school system for their results on another international test, PISA. However university and technical college mathematics faculty in Finland will not be surprised. See [this] article signed by over 200 of them.

Anybody who suggests the progressive/neoliberal education policies of the Scandinavian countries are worthy of emulation should be presented with these figures and laughed out of the room.

The results for individual American and Canadian states:

	A	B	C	D
Mass.	21.4	20.8	9.9	44.4
Calif.	28.2	21.6	11	38
Minn.	23.5	26.3	14	35.1
Quebec	27.3	23	13	35
Ontario	27.7	22.4	14	32.5
Conn.	21.8	25.8	17.7	31.3
Alberta	34.7	23.7	12.3	27.8

If you liked this article, you might also wish to check out:

The Geography Of Russia's IQ — Preliminary Thoughts on TIMSS/PIRLS 2012 — Education is the Elixir of Growth — Analysis Of Russia's PISA 2009 Results — Education is the Elixir of Growth III

Share this:

Comments

SH says:
July 22, 2013 at 3:39 pm

I agree that Sweden's educational system should only be considered as a deterrent. I am very surprised and puzzled by Finland equally poor performance on this particular task. After all, Finland have been touted as a great educational success story. Their PISA math scores, is only surpassed by East Asian countries. Teachers also have high status and salary in Finland, in complete contrast to the situation in Sweden. Despite awareness of the situation, I must say it is shocking only 1/7 of Swedish student could get this basic straightforward question right! Even a monkey would stand a 25% chance of picking the correct answer! US at 29% is also a shameful level in this regard.

charly says:
July 22, 2013 at 4:34 pm

This is a parsing problem but not a maths problem. How many 8 graders know that * means multiplication? My bet is that this number is very low in Sweden as they probably use the x for multiplications while it is commonly used in the far east

http://akarlin.com/2013/07/scandinavians-cant-do-fractions/

Note: Copy of the blogpost on the webpage: http://akarlin.com/2013/07/scandinavians-cant-do-fractions

Figure S1

Item M052228 from TIMSS 2011

Which shows a correct method for finding $\frac{1}{3} - \frac{1}{4}$?

A. $\frac{1 - 1}{4 - 3}$

B. $\frac{1}{4 - 3}$

C. $\frac{3 - 4}{3 \times 4}$

D. $\frac{4 - 3}{3 \times 4}$

Note: SOURCE item M052228: TIMSS 2011 Assessment. Copyright © 2013 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, Chestnut Hill, MA and International Association for the Evaluation of Educational Achievement (IEA), IEA Secretariat, Amsterdam, the Netherlands.

Figure S2

Item M032416 from TIMSS 2003 and TIMSS 2007

Which shows a correct procedure for finding $\frac{1}{5} - \frac{1}{3}$?

Ⓐ $\frac{1}{5} - \frac{1}{3} = \frac{1 - 1}{5 - 3}$

Ⓑ $\frac{1}{5} - \frac{1}{3} = \frac{1}{5 - 3}$

Ⓒ $\frac{1}{5} - \frac{1}{3} = \frac{5 - 3}{5 \times 3}$

Ⓓ $\frac{1}{5} - \frac{1}{3} = \frac{3 - 5}{5 \times 3}$

Note: SOURCE item M032416: TIMSS 2007 Assessment. Copyright © 2009 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Table S1

Overall Mathematics Score and Percent Correct Student Responses on Item M032416 in TIMSS 2003 (Martin, 2005) *and in TIMSS 2007* (Foy & Olson, 2009)*, and for Item M052228 in TIMSS 2011* (Foy et al., 2013) *for Selected Countries*

Cycle	**2003**		**2007**		**2011**	
Country	**Overall score (SE)**	**Item % correct (n)**	**Overall score (SE)**	**Item % correct (n)**	**Overall score (SE)**	**Item % correct (n)**
Norway	461 (2.5)	10.0 (344)	469 (2.0)	9.7 (660)	475 (2.4)	19.0 (538)
Sweden	499 (2.6)	9.3 (358)	491 (2.3)	7.6 (739)	484 (1.9)	14.4 (774)
Finland	-	-	-	-	514 (2.5)	16.1 (605)
Lithuania	502 (2.5)	39.0 (409)	506 (2.3)	33.5 (574)	502 (2.5)	28.5 (662)
England	498 (4.7)	10.6 (230)	513 (4.8)	17.0 (561)	507 (5.5)	28.2 (555)
Italy	484 (3.2)	42.1 (349)	480 (3.0)	47.0 (636)	498 (2.4)	50.9 (573)
Romania	475 (4.8)	39.7 (348)	461 (4.1)	37.4 (593)	458 (4.0)	35.7 (794)
Cyprus	459 (1.7)	15.0 (331)	465 (1.6)	15.5 (623)	-	-
Singapore	605 (3.6)	73.3 (507)	593 (3.8)	69.8 (649)	611 (3.8)	83.1 (850)
Korea	589 (2.2)	79.2 (437)	597 (2.7)	78.8 (604)	613 (2.9)	86.0 (739)
Botswana	366 (2.6)	17.5 (425)	364 (2.3)	19.0 (595)	397 (2.5)	21.6 (776)
South Africa	264 (5.5)	8.0 (746)	-	-	352 (2.5)	12.2 (1714)
International Average	*467*	*31.3*	*500*	*29.8*	*500*	*37.1*

Note: No similar item in the 2015 cycle. One similar item in 2019, but the data for this item is not available: «Item deleted for all countries, M10_3, ME72038, MP72038 (severe differential item functioning)» (Martin et al., 2020, p. 10_63). SE = standard error. n = number of students responding to the item.

Table S2

Overview and percent responses of first and second choice of distractors for focus item and its clones for students within selected countries (Foy et al., 2013; Foy & Olson, 2009; Martin, 2005).

Cycle	**2003**		**2007**		**2011**	
Country	**1st choice (%)**	**2nd choice (%)**	**1st choice (%)**	**2nd choice (%)**	**1st choice (%)**	**2nd choice (%)**
Norway	A (40.7)	B (30.7)	A (34.5)	B (30.7)	A (36.7)	B (26.4)
Sweden	B (39.3)	A (35.8)	A (37.7)	B (36.7)	A (36.7)	B (30.3)
Finland	-	-	-	-	A (42.3)	B (29.5)
Lithuania	B (37.9)	A (15.8)	B (40.9)	A (15.8)	B (37.2)	A (24.5)
England	A (38.1)	B (35.6)	B (33.9)	A (26.2)	B (32.8)	A (24.5)
Italy	B (21.9)	A (18.1)	B (27.5)	C (13.9)	B (23.0)	A (13.9)
Romania	B (32.5)	A (18.2)	B (32.9)	A (20.4)	B (30.6)	A (25.7)
Cyprus	B (41.6)	A (29.1)	B (36.6)	A (32.2)	-	-
Singapore	C (14.3)	A (6.2)	C (13.5)	A (8.2)	C (6.5)	B (5.5)
Korea	B (8.3)	C (7.6)	B (9.8)	C (8.4)	B (6.9)	C (4.2)
Botswana	B (34.9)	A (24.1)	B (31.2)	A (25.7)	A (37.0)	B (30.9)
South Africa	B (38.2)	A (32.2)	-	-	A (39.7)	B (35.6)
International	*B (29.3)*	*A (23.5)*	*B (29.2)*	*A (24.3)*	*B (26.0)*	*A (25.4)*

Note: The item M052228 («Which shows a correct method for finding ⅓ – ¼? ») and its clone M032416 («Which shows a correct method for finding ⅕ – ⅓? ») are multiple choice items with four response options each. The four response options have the same structure, in the same order, on both items. Response option A has 1-1 as numerator, and the denominator is the smaller denominator subtracted from the larger denominator of the two fractions in the question. Response option B is similar to the first response option, the only difference being that the numerator is set to 1. In response option C, the numerator is the second denominator subtracted from the first denominator, and the product of the two denominators from the question is the denominator. Response option D is the correct response option on both items.

Table S3

Descriptive IRT Non-Invariant Scaling across Countries to Check Item Easiness Ranking Differences between Countries for Norway, Sweden, Finland, and England

Country	Norway		Sweden		Finland		England	
Item[1]	Difficulty	SE	Difficulty	SE	Difficulty	SE	Difficulty	SE
ME72052	-2.78	0.18	-2.61	0.18	-1.69	0.12	-1.68	0.14
ME72002	-2.49	0.17	-2.30	0.18	-2.78	0.18	-0.23	0.12
ME72131	-2.49	0.16	-1.39	0.12	-2.29	0.15	-2.48	0.17
ME52087	-2.26	0.18	-2.67	0.22	-2.78	0.20	-2.10	0.17
ME62317	-2.21	0.16	-1.72	0.15	-1.29	0.12	-0.90	0.13
ME72109	-2.05	0.16	-1.99	0.16	-2.55	0.18	-1.73	0.16
ME62287	-2.03	0.15	-1.81	0.15	-2.12	0.15	-2.27	0.18
ME72120	-1.97	0.14	-2.01	0.15	-1.86	0.13	-1.52	0.14
ME72153B	-1.88	0.14	-1.44	0.14	-1.98	0.14	-1.86	0.15
ME62084	-1.77	0.13	-1.49	0.13	-1.43	0.11	-2.14	0.16
ME52146B	-1.76	0.14	-0.41	0.11	-3.13	0.21	-1.61	0.15
ME62192	-1.68	0.13	-2.65	0.19	-1.59	0.12	-1.95	0.15
ME52105	-1.67	0.13	-1.34	0.13	-1.54	0.12	-1.25	0.13
ME72041B	-1.61	0.12	-2.85	0.20	-0.52	0.10	0.97	0.12
ME72209	-1.61	0.14	-0.09	0.12	-1.58	0.13	-0.89	0.13
ME52068	-1.60	0.12	-1.31	0.13	-1.72	0.13	-2.06	0.15
ME52117	-1.60	0.12	-1.17	0.12	-1.79	0.12	-0.60	0.12
ME52125	-1.56	0.12	-2.04	0.14	0.32	0.09	0.74	0.11
ME62056	-1.51	0.13	-1.21	0.13	-1.82	0.13	-2.65	0.19
ME72017	-1.49	0.14	-1.36	0.14	-1.54	0.13	-1.08	0.14
ME52208	-1.49	0.12	-1.33	0.13	-1.71	0.13	-1.84	0.14
ME72220	-1.47	0.12	-1.49	0.13	-1.44	0.12	-1.43	0.14
ME72001	-1.43	0.14	-1.59	0.15	0.18	0.10	-0.58	0.14
ME72229	-1.42	0.13	-1.15	0.13	-1.94	0.14	-1.69	0.16
ME72206	-1.41	0.13	-0.67	0.12	-1.34	0.12	-1.45	0.14
ME72081	-1.40	0.12	-1.32	0.12	-1.67	0.12	-1.62	0.14
ME62254	-1.40	0.13	0.31	0.12	-1.16	0.11	-1.97	0.16
ME62261	-1.35	0.12	-1.39	0.13	-1.28	0.11	-1.85	0.15
ME72041A	-1.31	0.12	-2.48	0.17	-0.23	0.10	1.58	0.13
ME52130	-1.29	0.11	-0.86	0.11	-1.56	0.12	-1.17	0.13
ME72080	-1.26	0.12	-0.69	0.11	-0.86	0.10	-1.08	0.13
ME72059	-1.25	0.12	-0.56	0.11	-1.66	0.13	-0.21	0.12
ME72147	-1.23	0.12	-1.01	0.12	-1.18	0.11	-1.76	0.14
ME72074	-1.22	0.14	-1.13	0.14	-1.48	0.14	-0.99	0.15
ME72106C	-1.21	0.13	-0.28	0.11	-1.84	0.14	-1.01	0.13
ME72178	-1.15	0.11	-2.34	0.16	-1.42	0.11	-0.83	0.12
ME72198	-1.10	0.12	-1.94	0.15	-1.00	0.11	-0.14	0.12
ME62250B	-1.07	0.12	-0.39	0.11	-0.47	0.10	-1.22	0.14
ME52092	-1.06	0.11	-1.51	0.13	-1.12	0.11	-1.56	0.14
ME62230	-1.06	0.11	-0.78	0.11	-1.16	0.11	-0.52	0.12

Country	Norway		Sweden		Finland		England	
Item[1]	Difficulty	SE	Difficulty	SE	Difficulty	SE	Difficulty	SE
ME72108B	-1.04	0.13	-0.52	0.12	-0.90	0.12	0.04	0.12
ME52046	-0.99	0.11	-1.15	0.12	-0.47	0.10	-0.72	0.12
ME52110	-0.97	0.13	-0.95	0.13	-1.85	0.14	-0.76	0.13
ME72164	-0.92	0.11	-0.08	0.10	-1.20	0.11	-0.19	0.11
ME72090	-0.87	0.11	-0.80	0.12	-0.53	0.10	-0.44	0.12
ME72110B	-0.82	0.13	-1.94	0.16	-1.33	0.12	-1.20	0.15
ME72076	-0.79	0.11	-0.52	0.11	-0.92	0.10	0.21	0.11
ME52063	-0.77	0.10	-0.20	0.10	-0.57	0.10	0.07	0.11
ME62341	-0.76	0.11	-1.05	0.12	-0.40	0.10	-0.78	0.12
ME72098	-0.73	0.12	-0.03	0.11	-0.87	0.11	-1.02	0.14
ME62301	-0.67	0.11	-0.33	0.11	-1.50	0.12	-0.64	0.12
ME62350	-0.67	0.11	-1.20	0.12	-1.36	0.12	-1.31	0.13
ME72139	-0.66	0.11	-0.71	0.12	-0.48	0.10	-1.27	0.14
ME72013	-0.65	0.10	-1.35	0.13	-0.50	0.10	-0.04	0.11
ME52058B	-0.63	0.11	-0.41	0.11	-0.89	0.11	-0.86	0.13
ME72234	-0.62	0.10	-0.16	0.10	-0.59	0.10	-0.46	0.11
ME72035	-0.61	0.11	-1.18	0.13	-0.87	0.11	-1.07	0.14
ME72106B	-0.60	0.11	-0.15	0.11	-0.92	0.11	-1.08	0.13
ME72005	-0.55	0.10	-1.30	0.12	-1.39	0.11	0.81	0.11
ME52036	-0.53	0.10	-0.74	0.11	-0.67	0.10	-0.80	0.12
ME62288	-0.52	0.12	-1.07	0.15	-0.14	0.10	-0.65	0.14
ME72024	-0.52	0.11	-0.34	0.11	-0.54	0.10	-0.83	0.13
ME52048	-0.50	0.10	-0.57	0.11	-0.75	0.10	1.30	0.12
ME62173	-0.48	0.11	-0.93	0.12	-0.84	0.10	-0.97	0.12
ME62146	-0.46	0.10	-1.37	0.12	-0.52	0.09	0.12	0.11
ME62105	-0.46	0.11	-0.28	0.11	-0.77	0.11	-0.83	0.13
ME72007	-0.455	0.10	-0.80	0.11	-0.65	0.10	-1.22	0.13
ME62002	-0.44	0.10	-0.45	0.10	-0.22	0.09	-0.10	0.11
ME72056	-0.44	0.11	-0.65	0.12	-0.62	0.10	0.02	0.12
ME52229	-0.44	0.10	-0.06	0.10	0.16	0.092	-0.43	0.12
ME62115	-0.39	0.10	-0.47	0.11	-0.15	0.09	-0.41	0.12
ME72161	-0.37	0.10	-0.21	0.10	-0.52	0.10	-0.58	0.12
ME72125	-0.33	0.10	0.32	0.10	-0.04	0.09	-0.43	0.12
ME72110A	-0.32	0.11	-0.98	0.12	-0.68	0.10	-0.35	0.12
ME72126	-0.30	0.11	-0.04	0.11	-0.53	0.10	-0.78	0.13
ME62078	-0.29	0.11	0.30	0.11	-0.34	0.10	-0.66	0.13
ME62030	-0.27	0.10	0.56	0.10	0.33	0.09	0.10	0.12
ME72188	-0.27	0.10	-0.05	0.10	-0.07	0.09	-0.84	0.12
ME72181	-0.26	0.10	-0.34	0.11	0.22	0.09	-0.04	0.11
ME62143	-0.23	0.11	-0.90	0.13	-0.76	0.11	-0.08	0.13
ME62245	-0.22	0.10	-0.03	0.11	0.02	0.09	-0.03	0.11
ME52418A	-0.21	0.10	0.17	0.10	-0.11	0.09	-0.08	0.11
ME72083B	-0.15	0.10	-0.26	0.11	-0.04	0.09	-0.49	0.12
ME72108A	-0.15	0.10	-0.09	0.11	0.08	0.09	0.12	0.11
ME72103	-0.14	0.10	0.10	0.10	0.64	0.09	0.63	0.11
ME62346	-0.14	0.10	0.05	0.10	-0.55	0.10	-0.59	0.12

Country	Norway		Sweden		Finland		England	
Item[1]	Difficulty	SE	Difficulty	SE	Difficulty	SE	Difficulty	SE
ME52174B	-0.13	0.11	-0.32	0.11	-0.48	0.10	-1.06	0.14
ME62212	-0.10	0.10	0.11	0.10	0.09	0.09	-0.10	0.11
ME72021	-0.06	0.10	0.20	0.10	0.13	0.09	-0.45	0.12
ME62001	-0.04	0.10	-0.17	0.10	-0.30	0.09	-0.17	0.11
ME52073	-0.04	0.10	-0.08	0.10	0.10	0.09	-0.62	0.12
ME72187	-0.02	0.10	-0.31	0.11	0.84	0.09	0.60	0.11
ME62241	-0.01	0.11	-0.22	0.12	-0.42	0.10	-0.48	0.12
ME72095	-0.01	0.12	0.57	0.12	0.41	0.10	0.27	0.13
ME72221	-0.01	0.10	-0.34	0.11	0.29	0.09	-0.16	0.12
ME62284	0.00	0.10	-0.16	0.11	0.63	0.09	0.24	0.11
ME62151	0.01	0.10	0.14	0.10	0.19	0.09	-0.66	0.12
ME72026	0.03	0.09	0.16	0.10	0.59	0.09	-0.60	0.12
ME62149	0.03	0.10	0.16	0.10	0.56	0.09	-0.38	0.11
ME52147	0.05	0.10	0.02	0.10	0.16	0.09	0.46	0.11
ME72055	0.05	0.09	0.24	0.10	0.07	0.09	-0.62	0.112
ME72022	0.07	0.10	0.40	0.10	0.13	0.09	-0.30	0.11
ME72185A	0.07	0.10	-0.11	0.11	-0.67	0.10	-0.04	0.11
ME72045	0.08	0.10	-0.23	0.11	0.08	0.09	0.55	0.11
ME72043	0.10	0.10	-0.40	0.11	-0.15	0.09	-0.13	0.12
ME62174	0.11	0.10	0.23	0.10	0.44	0.09	0.10	0.11
ME72225	0.12	0.10	0.27	0.10	0.29	0.09	-0.04	0.12
ME62244	0.12	0.10	-0.45	0.12	0.19	0.09	-0.17	0.12
ME52039	0.14	0.10	-0.07	0.12	0.76	0.09	-0.31	0.12
ME72222	0.17	0.10	0.09	0.10	0.23	0.09	0.19	0.11
ME52413	0.18	0.09	-0.75	0.11	0.65	0.09	0.29	0.11
ME62027	0.18	0.10	0.52	0.10	0.10	0.09	0.41	0.12
ME72180	0.19	0.10	0.10	0.10	0.31	0.09	0.28	0.11
ME62005	0.19	0.09	0.25	0.10	0.17	0.09	-0.13	0.11
ME72185B	0.19	0.10	0.04	0.11	-0.55	0.10	0.16	0.12
ME72069	0.21	0.10	0.09	0.10	0.38	0.09	0.56	0.11
ME62139	0.24	0.10	0.26	0.10	0.09	0.09	-0.50	0.12
ME62344	0.24	0.10	-0.70	0.12	-2.14	0.15	-1.89	0.16
ME72196	0.24	0.11	-0.10	0.12	0.23	0.10	0.42	0.12
ME62120	0.27	0.10	0.74	0.10	0.40	0.09	0.03	0.11
ME62219	0.29	0.09	0.07	0.10	0.01	0.09	-0.07	0.11
ME62040	0.30	0.10	0.31	0.10	0.21	0.09	-0.17	0.11
ME62351	0.30	0.09	-0.25	0.10	-0.70	0.10	0.71	0.11
ME62320	0.32	0.10	0.51	0.10	0.39	0.09	0.02	0.12
ME52146A	0.32	0.10	0.49	0.10	-0.08	0.09	-0.28	0.12
ME72128B	0.32	0.11	0.66	0.11	0.67	0.10	-0.03	0.12
ME52067	0.34	0.10	-0.00	0.10	1.11	0.09	1.30	0.12
ME62095	0.34	0.10	0.84	0.10	0.51	0.09	0.10	0.11
ME52078	0.39	0.09	0.18	0.10	0.51	0.09	0.03	0.11
ME52083	0.39	0.10	0.43	0.10	0.12	0.09	-0.26	0.11
ME72128A	0.41	0.10	0.61	0.11	0.80	0.09	-0.14	0.12
ME72119	0.43	0.10	-0.27	0.11	0.45	0.09	0.21	0.11

Country	Norway		Sweden		Finland		England	
Item[1]	**Difficulty**	**SE**	**Difficulty**	**SE**	**Difficulty**	**SE**	**Difficulty**	**SE**
ME72223	0.43	0.09	0.77	0.10	0.40	0.09	0.45	0.11
ME72227	0.44	0.10	0.00	0.10	0.34	0.09	0.13	0.11
ME62171	0.47	0.10	-0.25	0.10	1.11	0.09	1.25	0.12
ME52421	0.49	0.10	1.63	0.12	0.22	0.09	0.27	0.11
ME62152	0.49	0.10	0.67	0.10	0.77	0.09	0.22	0.11
ME72067	0.51	0.10	-0.07	0.10	0.69	0.09	1.33	0.12
ME62300	0.53	0.10	1.01	0.10	0.76	0.09	0.47	0.11
ME52115	0.53	0.10	-0.01	0.10	-0.08	0.09	0.72	0.11
ME52418B	0.57	0.11	0.48	0.11	0.55	0.09	0.52	0.11
ME62271	0.58	0.10	0.79	0.10	0.51	0.09	0.01	0.11
ME72171	0.59	0.10	0.41	0.11	0.75	0.09	0.63	0.12
ME62133	0.59	0.10	0.33	0.10	0.52	0.09	0.22	0.11
ME62072	0.59	0.10	1.09	0.10	0.68	0.09	0.91	0.11
ME72189	0.60	0.10	0.78	0.10	0.55	0.09	0.05	0.12
ME62242	0.62	0.10	0.92	0.10	1.02	0.09	0.93	0.11
ME62132B	0.64	0.10	0.65	0.10	0.41	0.09	0.29	0.11
ME72019	0.64	0.10	0.95	0.10	0.81	0.09	0.44	0.11
ME52079	0.65	0.10	0.67	0.10	0.64	0.09	1.25	0.12
ME62170	0.68	0.10	0.80	0.10	1.10	0.10	0.44	0.12
ME62164	0.69	0.09	1.37	0.11	0.56	0.09	0.60	0.11
ME62214	0.72	0.10	0.72	0.10	0.68	0.09	-0.03	0.12
ME52215	0.80	0.10	0.29	0.10	1.50	0.10	1.06	0.11
ME72068	0.81	0.10	0.79	0.10	0.91	0.09	0.98	0.11
ME62250A	0.84	0.10	1.08	0.11	1.31	0.10	0.09	0.12
ME72172	0.86	0.10	1.22	0.11	1.01	0.09	0.79	0.11
ME72049	0.86	0.10	0.13	0.10	0.95	0.09	1.72	0.12
ME72025	0.90	0.10	1.09	0.10	0.66	0.09	0.35	0.11
ME62142	0.90	0.09	1.13	0.10	0.99	0.09	1.18	0.12
ME52034	0.91	0.10	0.99	0.10	1.03	0.09	0.49	0.11
ME52072	0.91	0.09	0.28	0.10	0.97	0.09	1.28	0.12
ME52174A	0.92	0.10	0.82	0.10	0.93	0.09	0.44	0.11
ME52082	0.94	0.10	1.02	0.10	0.81	0.09	0.77	0.11
ME62076	0.94	0.10	1.26	0.11	1.15	0.09	0.88	0.11
ME72106A	0.96	0.10	1.18	0.11	1.20	0.09	0.96	0.11
ME72233	0.98	0.10	0.99	0.10	1.26	0.09	0.42	0.11
ME72190	1.01	0.10	1.27	0.11	0.81	0.09	1.24	0.12
ME72140	1.05	0.10	0.94	0.10	0.78	0.09	1.03	0.11
ME62296	1.10	0.10	1.35	0.11	1.15	0.09	1.38	0.12
ME72153A	1.10	0.10	0.47	0.11	0.58	0.09	0.24	0.11
ME62223	1.10	0.10	1.76	0.11	1.43	0.09	1.19	0.12
ME72192	1.12	0.10	1.35	0.11	0.86	0.09	1.16	0.12
ME72232	1.13	0.10	0.81	0.11	1.03	0.09	1.48	0.12
ME52502	1.15	0.10	1.13	0.11	1.37	0.10	0.88	0.11
ME52024	1.18	0.10	0.93	0.10	0.92	0.09	0.44	0.11
ME72170	1.20	0.10	0.61	0.10	0.92	0.09	1.03	0.12
ME72094	1.21	0.10	1.86	0.11	1.20	0.09	1.13	0.12

Country	Norway		Sweden		Finland		England	
Item[1]	Difficulty	SE	Difficulty	SE	Difficulty	SE	Difficulty	SE
ME52364	1.24	0.10	1.36	0.11	2.01	0.11	2.05	0.14
ME62215	1.25	0.10	1.41	0.11	1.22	0.10	0.42	0.12
ME72154	1.28	0.10	1.55	0.11	1.64	0.10	0.99	0.11
ME62154	1.29	0.10	1.39	0.11	1.52	0.10	1.04	0.12
ME72121	1.32	0.10	0.85	0.10	1.08	0.09	1.57	0.13
ME72150	1.33	0.10	1.57	0.11	1.13	0.09	1.38	0.12
ME72083A	1.38	0.11	1.63	0.12	1.29	0.10	1.70	0.13
ME72211A	1.39	0.11	1.16	0.11	1.24	0.09	1.17	0.12
ME52204	1.45	0.10	0.40	0.10	-2.03	0.14	1.64	0.12
ME62067	1.47	0.10	1.46	0.11	1.61	0.10	1.27	0.12
ME52419A	1.56	0.10	1.26	0.11	1.57	0.10	1.17	0.12
ME52426	1.64	0.11	1.71	0.12	2.12	0.11	1.42	0.12
ME52407	1.64	0.11	0.97	0.10	1.17	0.09	0.55	0.11
ME62132A	1.69	0.11	1.52	0.12	1.77	0.10	1.16	0.12
ME52058A	1.70	0.11	1.62	0.11	2.11	0.11	1.51	0.12
ME72237	1.71	0.10	1.42	0.11	1.43	0.10	1.83	0.13
ME62335	1.71	0.11	0.76	0.10	1.02	0.10	1.58	0.12
ME72027	1.91	0.11	1.97	0.12	2.19	0.11	1.31	0.12
ME72020	2.01	0.11	1.33	0.10	2.36	0.11	1.57	0.12
ME62345A	2.07	0.09	1.07	0.08	0.89	0.07	1.90	0.10
ME52161	2.27	0.12	2.27	0.13	1.90	0.10	1.61	0.12
ME62194	2.31	0.12	2.10	0.12	2.22	0.11	1.51	0.12
ME62150	2.39	0.13	2.14	0.12	2.65	0.12	2.08	0.13
ME62329	2.47	0.12	2.41	0.13	2.65	0.12	2.24	0.14
ME52134	2.51	0.13	2.41	0.13	2.48	0.12	2.11	0.13
ME52419B	3.12	0.17	2.66	0.15	3.26	0.15	2.63	0.16

Note: The higher the difficulty estimate, the easier the task. The order is according to the Norwegian item ranks.

[1] The item names are color-coded where black is a mathematics task not involving fractions, blue is a fraction task, and red is a fraction task involving the use of a common denominator.

Table S4

Descriptive IRT Non-Invariant Scaling across Countries to Check Item Easiness Ranking Differences between Countries for Italy, Lithuania, Singapore, and Korea

Country	Italy		Lithuania		Singapore		Korea	
Item[1]	Difficulty	SE	Difficulty	SE	Difficulty	SE	Difficulty	SE
ME72052	-1.52	0.14	-1.72	0.13	-1.11	0.10	-1.28	0.11
ME72002	-0.05	0.11	-0.32	0.12	-0.84	0.10	0.17	0.12
ME72131	-1.61	0.14	-3.06	0.19	-1.25	0.10	-0.48	0.11
ME52087	-2.56	0.23	-2.53	0.19	-0.64	0.10	-0.07	0.12

Country	Italy		Lithuania		Singapore		Korea	
Item[1]	Difficulty	SE	Difficulty	SE	Difficulty	SE	Difficulty	SE
ME62317	-2.02	0.17	-1.05	0.12	-0.62	0.10	-1.65	0.11
ME72109	-1.57	0.16	-1.69	0.15	-1.29	0.10	-1.76	0.12
ME62287	-1.46	0.15	-2.18	0.16	-1.86	0.10	-1.65	0.11
ME72120	-2.24	0.18	-1.41	0.12	-0.12	0.10	-0.61	0.11
ME72153B	-2.12	0.19	-1.37	0.13	-0.44	0.10	-2.66	0.13
ME62084	-1.70	0.15	-1.18	0.12	-1.72	0.10	-0.61	0.11
ME52146B	-2.32	0.19	-2.42	0.17	-1.51	0.10	-2.08	0.12
ME62192	0.04	0.11	-1.12	0.12	-0.27	0.10	-2.57	0.12
ME52105	-2.90	0.22	-1.90	0.14	-2.10	0.10	-1.39	0.11
ME72041B	-0.66	0.12	-0.85	0.12	0.99	0.12	0.36	0.12
ME72209	-1.40	0.15	-1.86	0.15	-1.52	0.10	-2.89	0.13
ME52068	-1.07	0.13	-1.12	0.12	-0.94	0.10	-0.20	0.11
ME52117	-1.40	0.13	-2.18	0.15	-2.40	0.10	-3.26	0.13
ME52125	0.86	0.10	-1.20	0.12	0.11	0.11	-0.26	0.12
ME62056	-1.82	0.16	-1.50	0.14	-1.07	0.10	-0.59	0.11
ME72017	-1.76	0.17	-1.42	0.14	-1.39	0.10	-0.98	0.11
ME52208	-1.98	0.16	-1.20	0.12	-0.39	0.10	-1.49	0.11
ME72220	-0.78	0.12	-0.74	0.11	0.071	0.10	0.11	0.12
ME72001	-0.44	0.12	-0.33	0.12	-0.73	0.10	0.40	0.13
ME72229	-1.63	0.16	-1.68	0.15	-1.41	0.10	-2.17	0.12
ME72206	-1.76	0.16	-1.63	0.14	-1.72	0.10	-2.34	0.12
ME72081	-1.15	0.12	-1.34	0.12	-1.21	0.10	-0.60	0.11
ME62254	-1.58	0.17	-1.38	0.13	-0.79	0.10	-0.18	0.12
ME62261	-1.08	0.13	-2.05	0.14	-1.04	0.10	0.36	0.12
ME72041A	-0.15	0.11	-0.43	0.11	1.80	0.15	1.29	0.15
ME52130	-0.67	0.11	-0.38	0.11	0.089	0.10	0.14	0.12
ME72080	-0.47	0.11	-0.54	0.11	-0.03	0.10	-0.24	0.12
ME72059	-0.78	0.12	0.52	0.11	0.71	0.12	-0.03	0.12
ME72147	-1.08	0.13	-1.23	0.12	-1.08	0.10	-1.07	0.11
ME72074	-0.98	0.15	-0.53	0.13	-0.78	0.10	-0.63	0.11
ME72106C	-1.73	0.18	-1.67	0.15	-0.98	0.10	-1.23	0.11
ME72178	-0.20	0.11	-0.87	0.11	0.20	0.11	-1.71	0.11
ME72198	0.16	0.11	0.33	0.11	-0.26	0.10	0.05	0.12
ME62250B	-0.09	0.11	-0.33	0.11	-0.88	0.10	-0.44	0.11
ME52092	-0.93	0.12	-1.54	0.13	-2.39	0.10	-1.34	0.11
ME62230	-1.08	0.12	-1.25	0.12	0.04	0.11	-1.37	0.11
ME72108B	-0.50	0.13	-0.36	0.13	-0.72	0.10	-1.20	0.11
ME52046	-1.15	0.13	-0.99	0.12	-2.07	0.10	-2.32	0.11
ME52110	-0.10	0.12	-0.37	0.12	-0.31	0.10	-0.12	0.12
ME72164	-0.49	0.11	-1.06	0.12	-1.89	0.10	-1.57	0.11
ME72090	0.06	0.11	0.03	0.10	0.08	0.10	0.18	0.11
ME72110B	-0.98	0.14	-0.86	0.14	-1.24	0.10	-1.98	0.12
ME72076	-0.25	0.11	0.10	0.10	0.48	0.11	-0.07	0.12
ME52063	-0.31	0.11	0.12	0.10	1.20	0.13	0.59	0.13
ME62341	-0.87	0.12	-1.03	0.12	-1.25	0.10	-0.57	0.11
ME72098	-1.00627500	0.13	-0.18	0.11	0.14	0.11	-0.55	0.12

Country	Italy		Lithuania		Singapore		Korea	
Item[1]	Difficulty	SE	Difficulty	SE	Difficulty	SE	Difficulty	SE
ME62301	-0.88	0.12	-0.63	0.12	-1.52	0.10	-1.28	0.11
ME62350	-1.72	0.15	-2.05	0.15	-2.06	0.10	-1.59	0.11
ME72139	-1.60	0.15	-0.84	0.12	-0.88	0.10	-0.03	0.12
ME72013	0.18	0.10	0.41	0.10	0.13	0.11	-0.10	0.11
ME52058B	-1.02	0.13	-1.16	0.12	-0.89	0.10	-1.21	0.11
ME72234	0.26	0.10	-0.46	0.11	-0.85	0.10	-0.26	0.12
ME72035	0.49	0.11	0.04	0.11	-0.94	0.10	0.81	0.13
ME72106B	-0.96	0.13	-1.01	0.12	-0.62	0.10	-0.87	0.11
ME72005	-0.35	0.11	1.28	0.11	-0.23	0.10	1.56	0.15
ME52036	-0.21	0.11	-0.52	0.11	-0.81	0.10	-0.60	0.11
ME62288	-0.02	0.12	-0.35	0.12	-1.70	0.10	-1.22	0.10
ME72024	-0.09	0.12	-0.19	0.11	-0.37	0.10	-0.10	0.12
ME52048	0.038	0.11	0.51	0.10	-0.13	0.10	-1.56	0.11
ME62173	-0.56	0.11	-0.77	0.11	0.27	0.11	0.52	0.12
ME62146	-0.73	0.11	-0.13	0.10	0.68	0.11	0.75	0.13
ME62105	-0.61	0.12	-0.62	0.12	-0.07	0.10	-0.08	0.12
ME72007	-0.25	0.11	-0.49	0.11	-0.61	0.10	-0.38	0.11
ME62002	0.58	0.10	0.07	0.10	-0.82	0.10	-1.21	0.11
ME72056	0.21	0.11	0.47	0.11	0.12	0.11	-0.78	0.11
ME52229	0.23	0.10	0.40	0.11	0.11	0.11	0.92	0.14
ME62115	-0.62	0.11	-1.04	0.12	-1.90	0.10	-0.12	0.11
ME72161	-0.59	0.11	-0.54	0.11	-0.74	0.10	-0.66	0.11
ME72125	0.51	0.10	-0.05	0.10	0.37	0.11	-0.23	0.11
ME72110A	-0.20	0.11	-0.41	0.11	-0.64	0.10	-1.20	0.11
ME72126	0.13	0.12	-0.66	0.12	0.13	0.11	0.12	0.12
ME62078	-0.71	0.13	-0.14	0.11	-0.11	0.10	-0.19	0.11
ME62030	0.18	0.11	0.37	0.11	-0.33	0.10	-0.34	0.11
ME72188	0.50	0.10	0.10	0.10	-1.02	0.10	-0.57	0.11
ME72181	-1.54	0.14	-1.30	0.12	-2.88	0.11	-1.10	0.11
ME62143	-1.26	0.15	-0.81	0.13	0.59	0.12	-0.38	0.12
ME62245	-0.36	0.11	0.35	0.10	-0.90	0.10	-0.44	0.11
ME52418A	-0.43	0.11	0.09	0.10	0.18	0.11	-0.65	0.11
ME72083B	-0.09	0.11	0.15	0.10	-1.11	0.10	-0.36	0.11
ME72108A	1.05	0.11	1.11	0.11	-0.23	0.10	-0.30	0.12
ME72103	0.43	0.10	0.07	0.10	-0.47	0.10	-0.28	0.12
ME62346	-0.36	0.11	-0.57	0.11	-0.72	0.10	-1.36	0.11
ME52174B	-0.76	0.12	-0.73	0.12	-1.57	0.10	-0.84	0.11
ME62212	-0.75	0.12	-1.17	0.12	-1.07	0.10	-1.27	0.11
ME72021	0.07	0.11	-0.19	0.11	-0.22	0.10	0.43	0.13
ME62001	0.40	0.10	-0.11	0.10	-0.67	0.10	-0.52	0.11
ME52073	-0.13	0.10	0.05	0.10	0.22	0.11	0.57	0.13
ME72187	1.64	0.11	1.21	0.11	-0.00	0.11	0.89	0.13
ME62241	-0.28	0.13	-0.69	0.12	-0.37	0.10	0.25	0.12
ME72095	-0.33	0.13	0.11	0.12	-1.04	0.11	-0.01	0.12
ME72221	0.66	0.10	0.78	0.10	1.02	0.12	0.76	0.13
ME62284	1.58	0.11	1.46	0.11	-0.29	0.10	0.67	0.12

Country	Italy		Lithuania		Singapore		Korea	
Item[1]	Difficulty	SE	Difficulty	SE	Difficulty	SE	Difficulty	SE
ME62151	-0.44	0.11	-0.56	0.11	-0.81	0.10	-0.43	0.11
ME72026	-0.60	0.11	-0.22	0.10	-0.90	0.10	-0.28	0.11
ME62149	0.45	0.10	0.05	0.10	0.01	0.10	-1.19	0.11
ME52147	0.23	0.10	0.07	0.10	0.01	0.10	-0.21	0.11
ME72055	-1.83	0.15	-1.00	0.12	-1.43	0.10	-1.48	0.11
ME72022	-0.05	0.11	-0.22	0.11	-0.74	0.10	-0.90	0.11
ME72185A	-0.80	0.13	-1.43	0.13	-0.13	0.10	-0.10	0.12
ME72045	-0.62	0.12	0.11	0.11	0.63	0.12	0.74	0.13
ME72043	0.67	0.10	-0.38	0.11	1.69	0.14	0.71	0.13
ME62174	0.71	0.10	0.18	0.10	0.77	0.12	1.15	0.14
ME72225	-0.03	0.11	0.27	0.11	-0.07	0.10	0.13	0.12
ME62244	0.40	0.11	0.40	0.11	-0.70	0.10	-0.38	0.12
ME52039	0.32	0.11	0.17	0.11	0.57	0.11	1.68	0.16
ME72222	1.48	0.11	0.55	0.10	0.41	0.11	-1.11	0.11
ME52413	0.89	0.10	0.86	0.10	0.19	0.11	0.59	0.13
ME62027	-0.87	0.12	-0.77	0.11	-0.91	0.10	-0.88	0.11
ME72180	1.17	0.11	0.77	0.10	-0.16	0.10	-0.31	0.12
ME62005	0.89	0.10	0.46	0.10	-0.19	0.10	-0.15	0.12
ME72185B	-0.61	0.13	-1.16	0.12	0.19	0.11	0.03	0.12
ME72069	0.29	0.11	0.92	0.10	0.77	0.12	0.99	0.13
ME62139	-0.43	0.11	-0.15	0.11	-0.49	0.10	-1.24	0.11
ME62344	-1.31	0.14	-0.27	0.12	-0.71	0.10	-1.03	0.11
ME72196	0.04	0.12	0.063	0.11	-0.12	0.11	-1.07	0.11
ME62120	0.39	0.10	-0.20	0.11	-0.20	0.10	0.49	0.12
ME62219	-0.18	0.10	-0.29	0.11	-0.18	0.10	0.28	0.12
ME62040	0.51	0.10	0.15	0.10	-0.90	0.10	-0.51	0.11
ME62351	-0.79	0.12	-1.80	0.13	-0.16	0.10	-1.34	0.11
ME62320	-0.55	0.11	0.06	0.11	0.32	0.11	0.17	0.12
ME52146A	0.37	0.10	0.02	0.11	-0.02	0.10	-0.14	0.12
ME72128B	1.07	0.11	-0.07	0.11	0.17	0.11	0.24	0.12
ME52067	1.55	0.11	1.17	0.11	1.04	0.12	1.01	0.13
ME62095	0.23	0.11	0.44	0.10	0.07	0.11	-0.30	0.11
ME52078	-0.04	0.10	0.22	0.10	0.28	0.11	-0.67	0.11
ME52083	0.17	0.11	-0.15	0.11	0.12	0.11	-0.14	0.12
ME72128A	0.46	0.11	-0.58	0.11	-0.41	0.10	0.38	0.12
ME72119	0.56	0.10	0.15	0.10	0.06	0.10	1.35	0.14
ME72223	0.41	0.10	0.57	0.10	0.23	0.11	0.56	0.13
ME72227	0.02	0.11	0.34	0.10	0.08	0.10	-0.27	0.12
ME62171	0.65	0.10	0.96	0.10	0.11	0.11	1.58	0.15
ME52421	-0.53	0.12	-0.41	0.11	-0.20	0.10	-0.20	0.12
ME62152	-0.20	0.11	0.28	0.11	-0.58	0.10	-0.57	0.11
ME72067	0.21	0.10	1.30	0.11	1.41	0.13	1.26	0.15
ME62300	1.04	0.10	0.82	0.11	0.35	0.11	0.39	0.12
ME52115	0.18	0.10	0.34	0.10	1.13	0.13	-0.05	0.12
ME52418B	0.73	0.11	0.60	0.11	0.65	0.12	-0.23	0.12
ME62271	0.01	0.10	0.15	0.11	0.55	0.11	0.75	0.13

Country	Italy		Lithuania		Singapore		Korea	
Item[1]	Difficulty	SE	Difficulty	SE	Difficulty	SE	Difficulty	SE
ME72171	0.25	0.11	-0.08	0.12	0.67	0.11	0.73	0.13
ME62133	0.56	0.10	0.17	0.10	-0.13	0.10	0.08	0.12
ME62072	0.72	0.10	0.54	0.10	0.38	0.11	-0.21	0.11
ME72189	0.67	0.10	0.48	0.10	0.81	0.12	0.83	0.13
ME62242	0.94	0.10	0.66	0.11	0.48	0.11	0.17	0.12
ME62132B	0.43	0.10	0.29	0.10	-0.65	0.10	-0.16	0.12
ME72019	-0.62	0.11	0.04	0.11	-0.08	0.10	-0.48	0.11
ME52079	0.49	0.10	0.90	0.10	1.19	0.13	0.10	0.13
ME62170	1.29	0.11	1.25	0.11	-0.29	0.10	0.69	0.13
ME62164	0.73	0.10	0.64	0.10	0.65	0.11	0.11	0.12
ME62214	-0.44	0.11	0.11	0.11	-0.02	0.10	-0.52	0.11
ME52215	0.94	0.10	0.66	0.10	0.78	0.12	1.22	0.14
ME72068	0.79	0.10	1.43	0.11	1.54	0.14	1.34	0.15
ME62250A	1.03	0.11	1.02	0.11	0.09	0.11	0.18	0.12
ME72172	0.86	0.11	0.59	0.10	0.71	0.12	1.08	0.13
ME72049	0.79	0.10	1.34	0.11	2.11	0.16	1.91	0.16
ME72025	-0.11	0.11	-0.20	0.11	0.74	0.12	0.08	0.12
ME62142	0.87	0.10	1.19	0.11	0.99	0.12	0.45	0.13
ME52034	0.41	0.10	0.48	0.10	0.74	0.12	-0.04	0.12
ME52072	1.65	0.11	0.98	0.11	0.78	0.12	0.95	0.14
ME52174A	0.52	0.10	0.60	0.11	-0.15	0.10	0.23	0.12
ME52082	1.05	0.10	0.81	0.10	0.98	0.12	1.47	0.15
ME62076	1.43	0.11	1.37	0.11	1.76	0.15	1.16	0.14
ME72106A	0.82	0.11	0.88	0.11	0.73	0.11	0.47	0.12
ME72233	0.88	0.10	0.73	0.10	-0.34	0.10	0.10	0.12
ME72190	0.78	0.11	0.81	0.11	1.38	0.13	0.75	0.13
ME72140	0.34	0.10	1.09	0.11	-0.37	0.10	-0.08	0.12
ME62296	1.17	0.11	0.39	0.11	0.41	0.11	0.60	0.13
ME72153A	0.33	0.11	0.44	0.10	0.57	0.11	0.84	0.13
ME62223	1.57	0.11	1.53	0.11	1.60	0.14	1.21	0.14
ME72192	0.26	0.10	0.79	0.11	0.45	0.11	0.99	0.14
ME72232	1.03	0.11	0.84	0.10	0.57	0.11	0.54	0.12
ME52502	1.58	0.11	1.48	0.12	0.88	0.12	0.86	0.14
ME52024	0.41	0.10	0.57	0.10	0.19	0.11	0.75	0.13
ME72170	0.76	0.11	0.77	0.11	0.38	0.11	-0.06	0.12
ME72094	1.37	0.11	1.62	0.12	0.75	0.12	0.79	0.13
ME52364	1.91	0.11	1.54	0.11	0.90	0.12	1.56	0.15
ME62215	-0.14	0.11	0.54	0.11	0.56	0.12	-0.82	0.11
ME72154	1.37	0.11	1.57	0.11	1.17	0.13	0.55	0.13
ME62154	0.70	0.10	1.39	0.11	0.76	0.12	1.33	0.14
ME72121	1.06	0.11	1.42	0.11	1.67	0.14	1.41	0.15
ME72150	1.04	0.10	1.53	0.11	1.55	0.13	1.01	0.13
ME72083A	1.72	0.12	1.34	0.12	0.91	0.12	1.19	0.15
ME72211A	0.83	0.11	1.36	0.11	1.44	0.13	1.18	0.14
ME52204	0.96	0.10	-1.05	0.12	0.55	0.11	1.11	0.14
ME62067	1.07	0.10	1.15	0.11	0.94	0.12	1.36	0.14

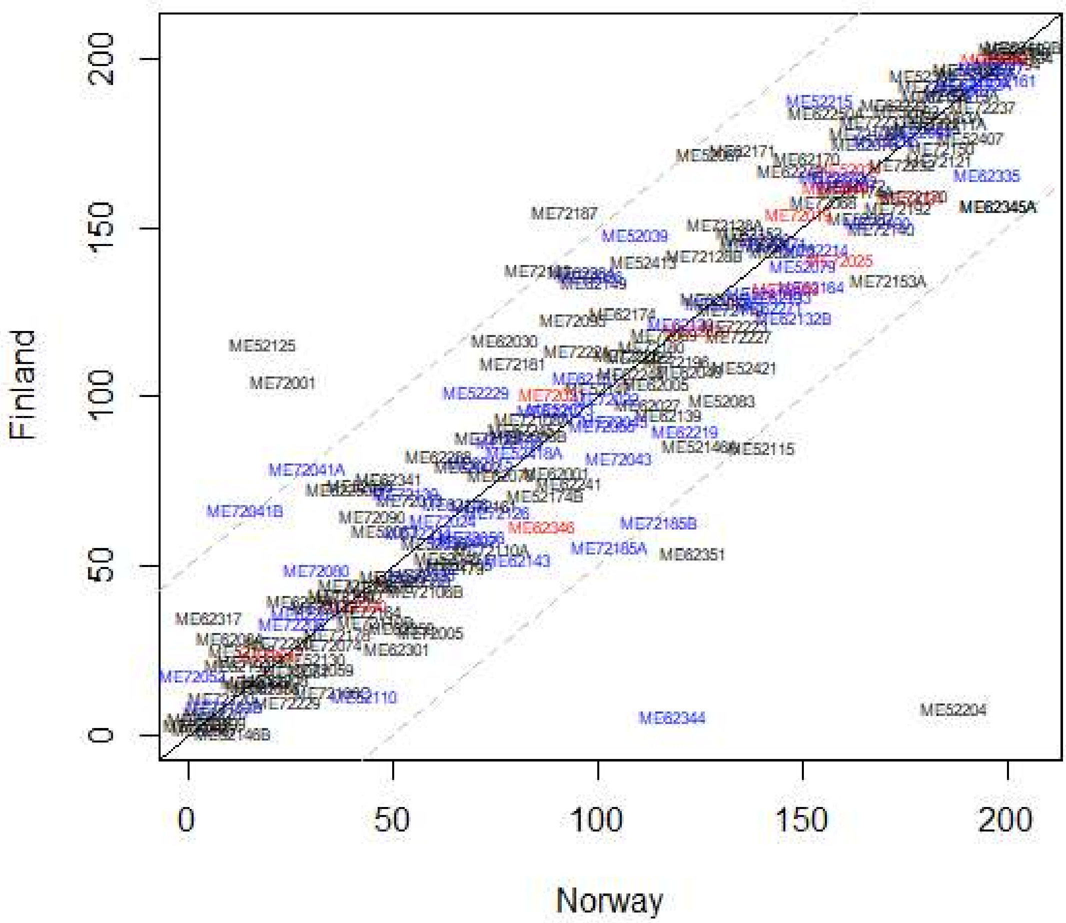
Country	Italy		Lithuania		Singapore		Korea	
Item[1]	Difficulty	SE	Difficulty	SE	Difficulty	SE	Difficulty	SE
ME52419A	1.02	0.10	1.53	0.11	0.39	0.11	0.40	0.12
ME52426	1.93	0.12	2.03	0.12	1.09	0.12	1.63	0.16
ME52407	1.08	0.10	1.19	0.11	0.66	0.11	1.72	0.16
ME62132A	1.35	0.11	1.52	0.11	0.81	0.12	1.01	0.14
ME52058A	1.39	0.11	1.54	0.11	1.49	0.14	1.35	0.15
ME72237	0.88	0.10	1.65	0.11	0.74	0.12	0.90	0.13
ME62335	1.49	0.11	1.00	0.10	1.89	0.15	2.03	0.17
ME72027	0.65	0.10	0.83	0.10	1.13	0.13	0.25	0.12
ME72020	1.18	0.10	1.25	0.11	0.85	0.12	0.89	0.14
ME62345A	0.96	0.08	1.05	0.08	0.99	0.09	0.13	0.09
ME52161	1.63	0.11	1.75	0.11	1.04	0.12	1.47	0.15
ME62194	1.84	0.11	2.33	0.13	1.57	0.14	2.04	0.17
ME62150	0.98	0.10	1.28	0.11	0.64	0.11	0.70	0.13
ME62329	2.41	0.13	2.07	0.12	1.63	0.14	1.01	0.13
ME52134	1.49	0.11	1.50	0.11	0.84	0.12	1.20	0.14
ME52419B	1.79	0.11	2.60	0.14	1.69	0.15	2.24	0.18

Note: The higher the difficulty estimate, the easier the task. The order is according to the Norwegian item ranks.

[1] The item names are color-coded where black is a mathematics task not involving fractions, blue is a fraction task, and red is a fraction task involving the use of a common denominator.

Figure S3

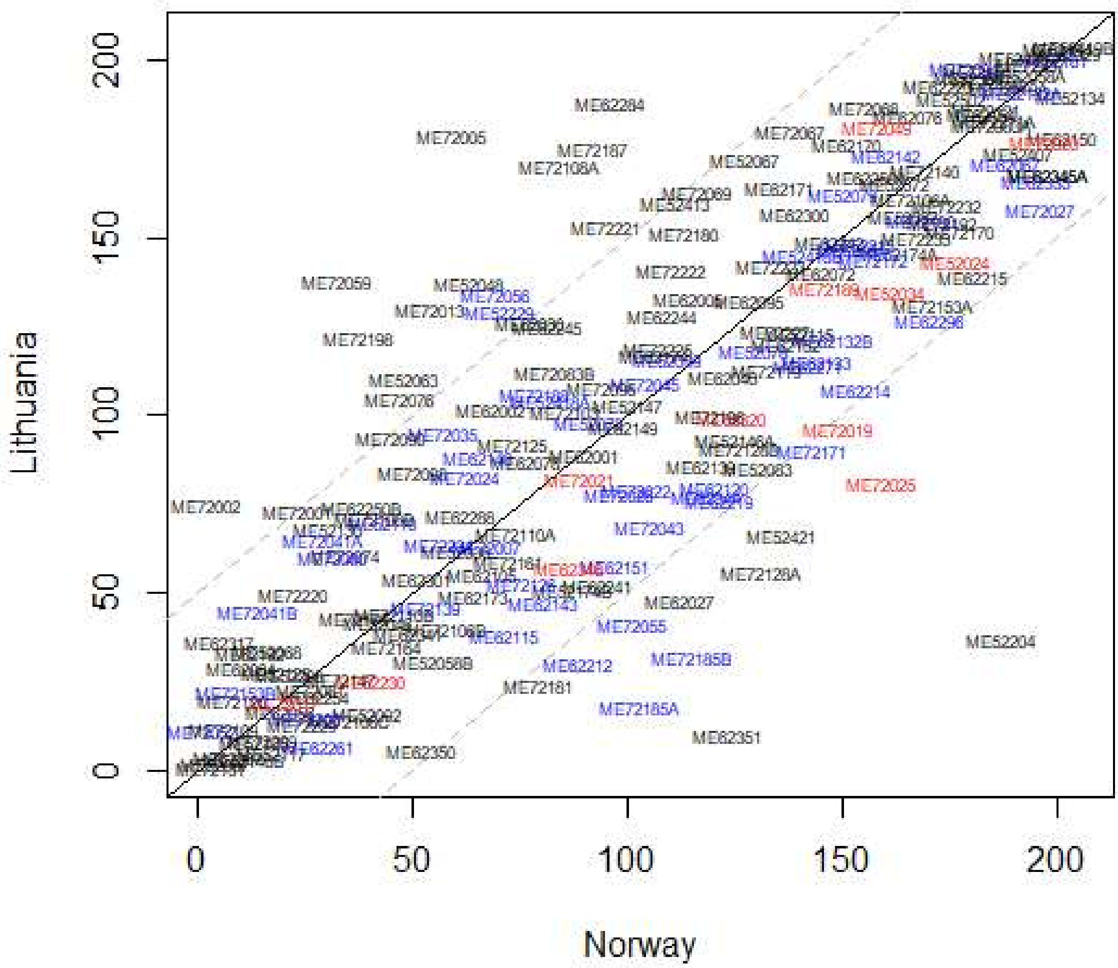
Landscape Plot Providing Pairwise Comparions to Check Item Easiness Ranking Differences between Norway and Finland (Nordic Country)



Note: The scatterplot represents item easiness for the same sample of items under an item selection rule for Norway (horistontal axis) and an item selection rule for Finland (vertical axis). Each code represents the pairwise combination of item easiness for one specific item. The items are color-coded so that items in mathematics are grey, fraction items are blue, and common denominator items are red. The solid black diagonal line with intercept 0 and slope 1 divides the plot: Codes on the line have the same item easiness for both countries. The further away an item is from the diagonal, the larger the difference in relative easiness ranking of the item between the two countries compared. For reference, the dashed grey diagonal lines indicate +/- 50 ranks different. The landscape plot is reflecting a similar cross-country correspondence in relative easiness of the items.

Figure S4

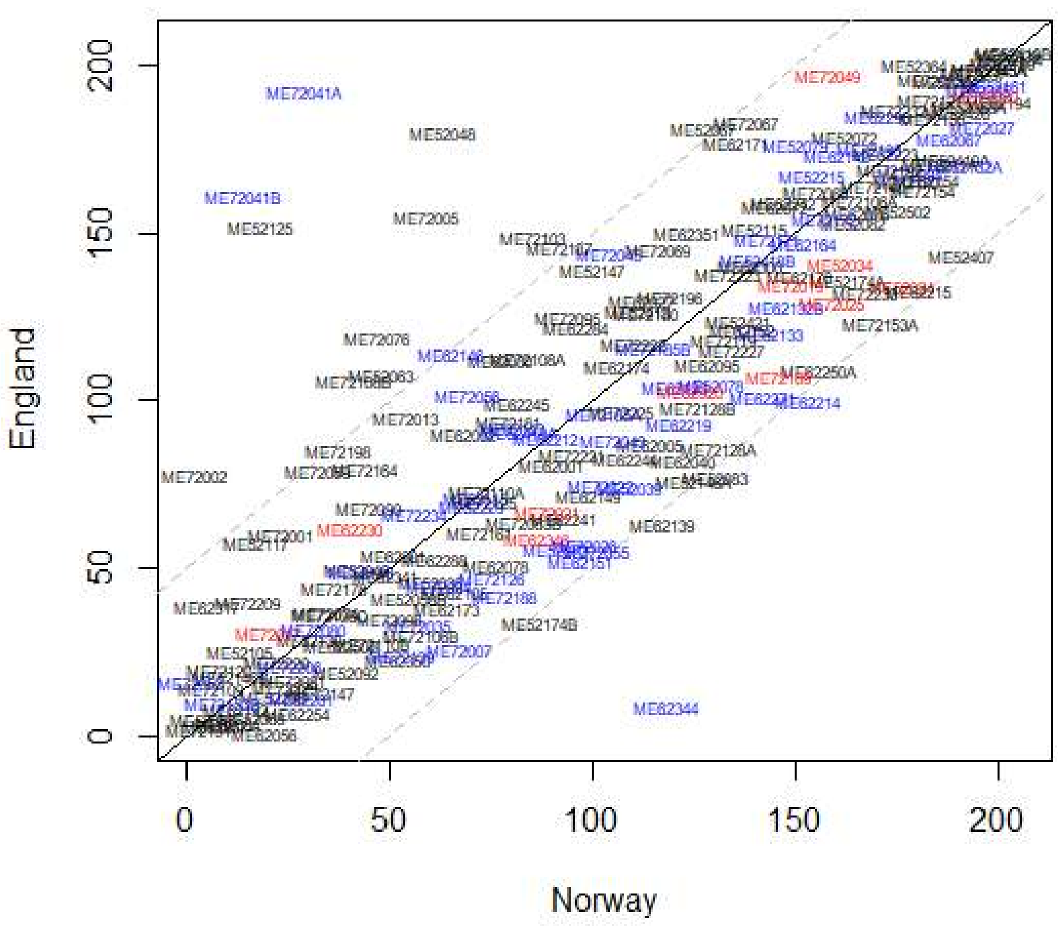
Landscape Plot Providing Pairwise Comparions to Check Item Easiness Ranking Differences between Norway and Lithuania (Baltic Country)



Note: The scatterplot represents item easiness for the same sample of items under an item selection rule for Norway (horistontal axis) and an item selection rule for Lithuania (vertical axis). Each code represents the pairwise combination of item easiness for one specific item. The items are color-coded so that items in mathematics are grey, fraction items are blue, and common denominator items are red. The solid black diagonal line with intercept 0 and slope 1 divides the plot: Codes on the line have the same item easiness for both countries. The further away an item is from the diagonal, the larger the difference in relative easiness ranking of the item between the two countries compared. For reference, the dashed grey diagonal lines indicate +/- 50 ranks different. The landscape plot is reflecting a fairly similar cross-country correspondence in relative easiness of the items.

Figure S5

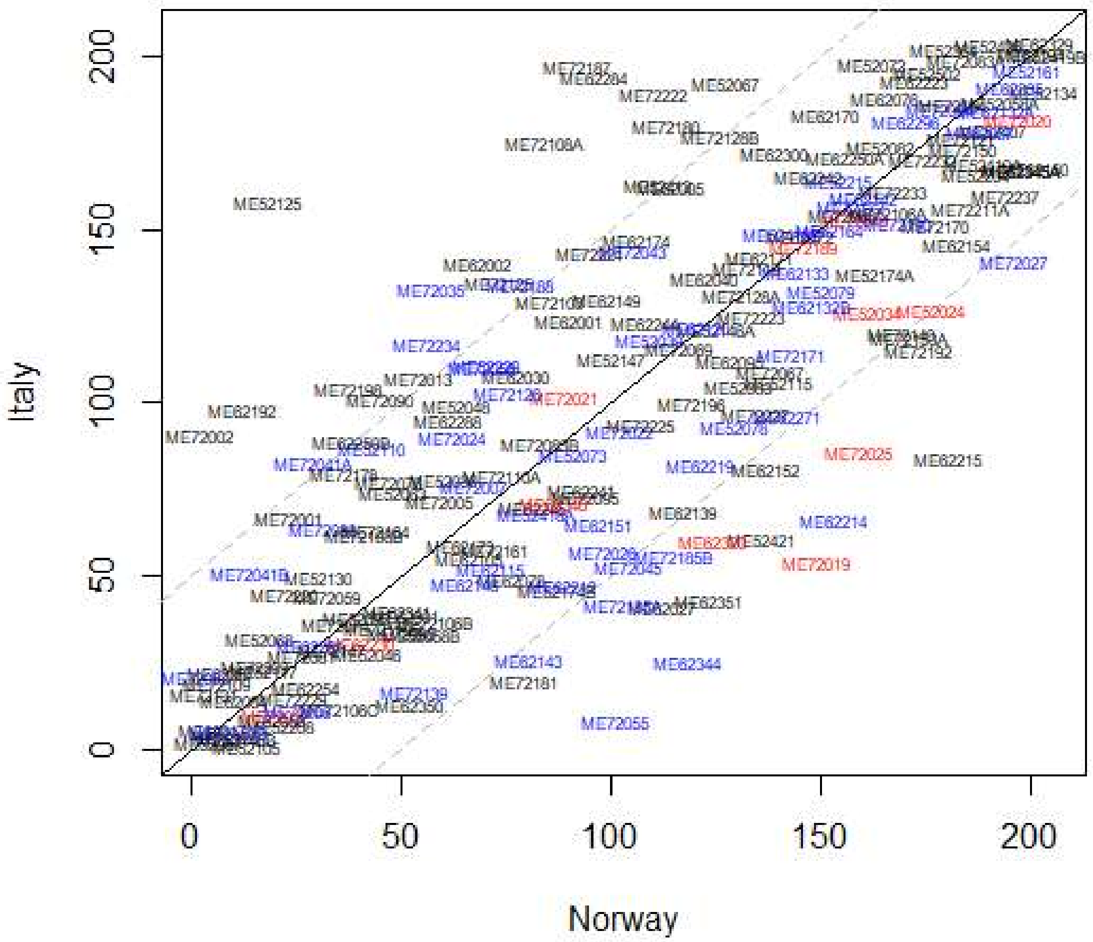
Landscape Plot Providing Pairwise Comparions to Check Item Easiness Ranking Differences between Norway and England (European Country)



Note: The scatterplot represents item easiness for the same sample of items under an item selection rule for Norway (horistontal axis) and an item selection rule for England (vertical axis). Each code represents the pairwise combination of item easiness for one specific item. The items are color-coded so that items in mathematics are grey, fraction items are blue, and common denominator items are red. The solid black diagonal line with intercept 0 and slope 1 divides the plot: Codes on the line have the same item easiness for both countries. The further away an item is from the diagonal, the larger the difference in relative easiness ranking of the item between the two countries compared. For reference, the dashed grey diagonal lines indicate +/- 50 ranks different. The landscape plot is reflecting a similar cross-country correspondence in relative easiness of the items.

Figure S6

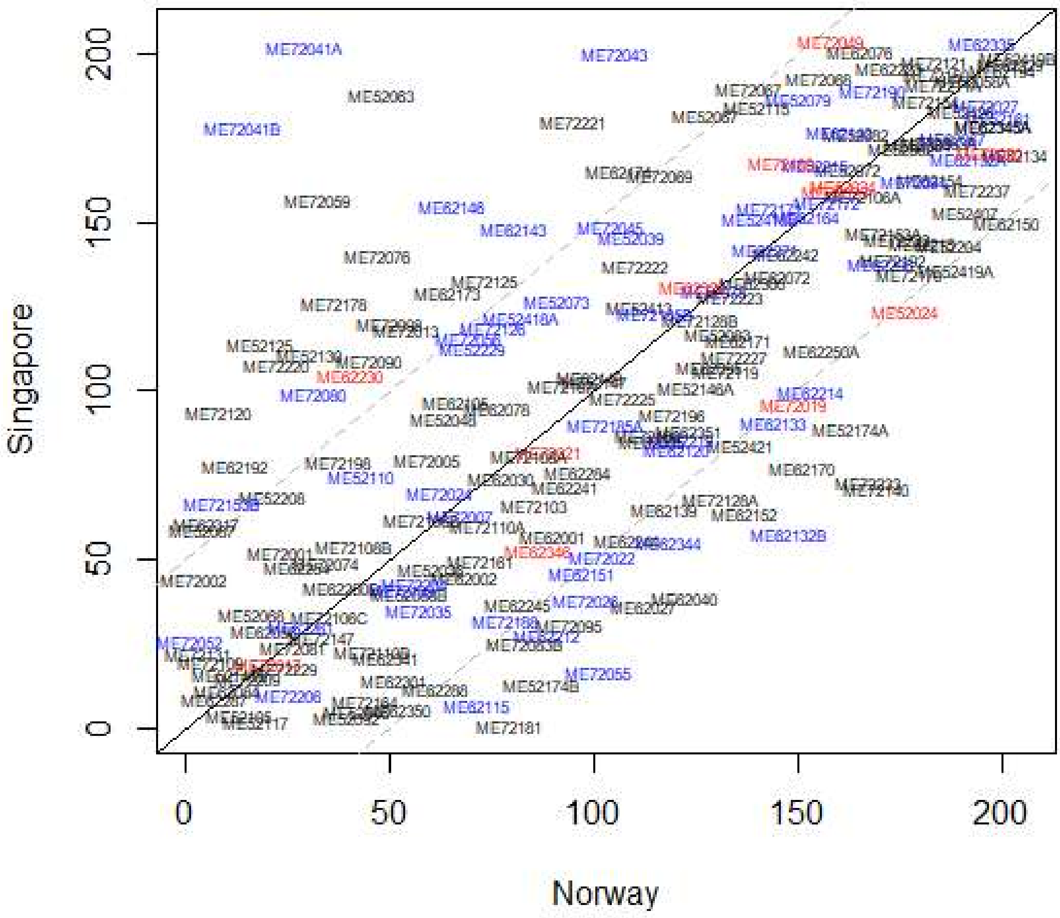
Landscape Plot Providing Pairwise Comparions to Check Item Easiness Ranking Differences between Norway and Italy (European Country)



Note: The scatterplot represents item easiness for the same sample of items under an item selection rule for Norway (horistontal axis) and an item selection rule for Italy (vertical axis). Each code represents the pairwise combination of item easiness for one specific item. The items are color-coded so that items in mathematics are grey, fraction items are blue, and common denominator items are red. The solid black diagonal line with intercept 0 and slope 1 divides the plot: Codes on the line have the same item easiness for both countries. The further away an item is from the diagonal, the larger the difference in relative easiness ranking of the item between the two countries compared. For reference, the dashed grey diagonal lines indicate +/- 50 ranks different. The landscape plot is reflecting a fairly similar cross-country correspondence in relative easiness of the items.

Figure S7

Landscape Plot Providing Pairwise Comparions to Check Item Easiness Ranking Differences between Norway and Singapore (High Achieving Country)



Note: The scatterplot represents item easiness for the same sample of items under an item selection rule for Norway (horistontal axis) and an item selection rule for Singapore (vertical axis). Each code represents the pairwise combination of item easiness for one specific item. The items are color-coded so that items in mathematics are grey, fraction items are blue, and common denominator items are red. The solid black diagonal line with intercept 0 and slope 1 divides the plot: Codes on the line have the same item easiness for both countries. The further away an item is from the diagonal, the larger the difference in relative easiness ranking of the item between the two countries compared. For reference, the dashed grey diagonal lines indicate +/- 50 ranks different. The landscape plot is reflecting a lower cross-country correspondence in relative easiness of the items.