

# Towards validation of a rational number instrument: An application of Rasch measurement theory



## Authors:

Jeremiah Maseko<sup>1</sup>   
Kakoma Luneta<sup>1</sup>   
Caroline Long<sup>1</sup>

## Affiliations:

<sup>1</sup>Department of Childhood Education, Faculty of Education, University of Johannesburg, Johannesburg, South Africa

## Corresponding author:

Jeremiah Maseko,  
jerrym@uj.ac.za

## Dates:

Received: 01 Aug. 2018  
Accepted: 23 Sept. 2019  
Published: 05 Dec. 2019

## How to cite this article:

Maseko, J., Luneta, K., & Long, C. (2019). Towards validation of a rational number instrument: An application of Rasch measurement theory. *Pythagoras*, 40(1), a441. <https://doi.org/10.4102/pythagoras.v40i1.441>

## Copyright:

© 2019. The Authors.  
Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

## Read online:



Scan this QR code with your smart phone or mobile device to read online.

The rational number knowledge of student teachers, in particular the equivalence of fractions, decimals, and percentages, and their comparison and ordering, is the focus of this article. An instrument comprising multiple choice, short answer and constructed response formats was designed to test conceptual and procedural understanding. Application of the Rasch model enables verification of whether the test content was consistent with the construct under investigation. The validation process was enabled by making explicit the expected responses according to the model versus actual responses by the students. The article shows where the Rasch model highlighted items that were consistent with the model and those that were not. Insights into both the construct and the instrument were gained. The test items showed good fit to the model; however, response dependency and high residual correlation within sets of items was detected. Strategies for resolving these issues are discussed in this article. We sought to answer the research question: to what extent does this test instrument provide valid information that can be used to inform teaching and learning of fractions? We were able to conclude that a refined instrument applied to first-year students at university provides useful information that can inform the teaching and learning of rational number concepts, a concept that runs through mathematics curricula from primary to university. Previously, most research on rational number concept has been conducted on young learners at school.

**Keywords:** fractions; equivalence; decimals; Rasch model; teacher education; percentage; compare; conversion.

## Introduction

Venkat and Spaul (2015) reported that 79% of 401 South African Grade 6 mathematics teachers showed proficiency of content knowledge below Grade 6–7 level in a Southern and East African Consortium for Monitoring Educational Quality (SACMEQ) 2007 mathematics teacher test. Universities recruit and receive students from some of these schools where these teachers are teaching. In the previous years of teaching first-year students in the mathematics module in the Foundation Phase teacher development programme, we noticed that each cohort of prospective teachers come with knowledge bases that are at different levels. These classes, of students' with varied mathematics knowledge, are difficult to teach unless you have some idea of their conceptual and procedural gaps. This varied knowledge base is greatly magnified in the domain of rational numbers in which they are expected to be knowledgeable and confident in order to teach and lay a good foundation in future teaching. An instrument, functioning as a diagnostic and baseline test for the 2015 first-year Foundation Phase cohort, was constructed at the university level in the fractions-decimals-percentages triad. This instrument aimed at gauging the level of students' cognitive understanding of rational numbers as well as evaluating the validity of the instrument that was used to elicit their mathematical cognition. All the participants admitted into the Foundation Phase teacher training programme were tested on 93 items comprising multiple choice, short answer and constructed response formats. That elicited both conceptual and procedural understanding.

Application of the Rasch model enabled a finer analysis of the test construct, the individual item and person measures, and the overall test functioning through making explicit the expected responses according to the model versus the actual responses by the students. In addition, the test as a whole was investigated for properties that are requirements of valid measurement such as the local independence where each item functions independently of each of the other items.

This article reports on how students displayed gaps in their rational number knowledge base but focuses primarily on the validation of the instrument. The following questions are answered:

- To what extent does the test provide valid measures of student proficiency?
- How might the test be improved for greater efficiency of administration, and greater validity for estimating student proficiency?

The aims of the immediate analyses were to:

- Evaluate the assessment tool in terms of fit to the model, both item and person fit, thereby checking whether the tool was appropriate for this student cohort.
- Provide detailed descriptions of selected items in relation to the students taking the test.

The validity and reliability of the assessment tool were analysed through the Rasch model incorporating both the dichotomous and partial credit model using Rasch Unidimensional Measurement Models (RUMM) software (see Andrich, Sheridan, & Luo, 2013). The processes of analysis and refinement, and the final outcome of this cycle are described. As this test was used as a preliminary diagnostic instrument, we regard ongoing cycles of refinement as pertinent in the interests of informing the teaching of mathematics on fractions-decimals-percentages to preservice cohorts of teachers in our programme.

## Literature review

In an attempt to clarify the assessment, how it was conducted and its purpose, we provide the justification for the exercise. It was critical to ensure that certain conditions were satisfied in order to safeguard the effectiveness of the assessment as well as the validity of the test items. Stiggins and Chappuis (2005) explained that assessment must be guided by a clear purpose and it must accurately reflect the learning expectations. Wiliam (2011) affirms that a method of assessment must be capable of reflecting the intended target and also act as a tool for gauging teaching proficiency. These were the core intentions of the assessment in this research, and therefore the validation of the test as a whole, and the validation of independent items was critical and appropriate.

The learning and teaching of rational number concepts is particularly complex. The representation of a fraction  $\frac{6}{25}$  as  $\frac{6}{25} = 0.24$  has a meaning different to whole numbers 6 and 25. The numbers 6 and 25 are called local values while together, as a single entity, yielding 0.24, they constitute a global value and have a different meaning and value from 6 and 25 represented separately (Gabriel, Szucs, & Content, 2013; Sangwin, 2007). These authors found that it was not a simple process for either learners or adults to cross the bridge from whole numbers to fractions (global value form). Vamvakoussi and Vosniadou (2007, 2010) identify two distinguishing features of rational numbers. These are firstly that for each point on a number line, for example  $\frac{1}{2}$ , there are infinite representations, that is,  $\frac{2}{4}, \frac{3}{6}, \frac{4}{8}$  etc. Secondly, between any two points on the number line there are infinitely many

numbers. Given this complexity, the operations on rational numbers, for example addition, subtraction, multiplication and division, require procedures that may previously have been learned when working with natural numbers but that now appear to generate misconceptions and associated errors (Harvey, 2011; Pantziara & Philippou, 2012; Shalem, Smith & Sorto, 2014). In fact the operations on rational numbers are somewhat distinct, and require additional conceptual understanding together with the associated procedures.

Besides the features mentioned previously, there are different representational systems for rational numbers, namely common fractions, decimal fractions and percentages. While there is equivalence across the three systems within the triad fractions-decimals-percentages, this equivalence is not obvious at face value unless the student has understood the organising principles of each system. For instance, the denominator of a percentage representation is always 100, for common fractions the choice of denominator is infinite, while for decimal fractions, the denominator is 1 (one).

The apparent simplicity of the percentage because of its everyday use belies the complexity of this 'privileged proportion' (Parker & Leinhardt, 1995, p. 421). For example, an additive difference between two percentages, may be confused with a ratio difference.

Hiebert and Lefevre (1986, p. 3–4) define conceptual knowledge as 'knowledge that is rich in relationship, that can be thought of as a connected web of knowledge, a network in which linking relationships are as prominent as the discrete pieces of information'. Such knowledge is described as that which is interconnected through relationships at various levels of abstraction. Conceptual knowledge is essential for learners to have conceptual understanding as in its absence they will indulge ineffectively in problem solving and follow wrong procedures to solve them. Conceptual knowledge plays a more important role, although interactively the two facets support a solid knowledge foundation.

Stacey et al. (2001) found that preservice primary school teachers had problems understanding the size of decimals in relation to zero including limited awareness on the misconception that 'shorter is larger' among learners. Ryan and Williams (2007) also highlight and explain misconceptions and associated errors on adding and subtraction fractions, working with decimals, and the meaning of place value that are commonly committed by learners, such as having problems with zero when subtracting smaller from larger digits. Huang, Liu and Lin (2009) report that preservice teachers in Taiwan displayed better fraction knowledge of procedures but lacked conceptual knowledge because of the way they had received this knowledge themselves. They recommended that these preservice teachers need more opportunities to construct their conceptual knowledge before they graduate. Pesek, Gray and Golding (1997) believe that clear understanding of rational numbers is one of the most foundational sections in the primary school curriculum and

yet, presently, is one of the least understood by both teachers and learners. Identifying mathematical competence levels of incoming preservice teachers provides an opportunity for the timely remediation of at-risk students. The conceptual complexities that generate misconceptions and associated errors emerge from lack of conceptual understanding (Ryan & Williams, 2007; Charalambous & Pitta-Pantazi, 2005, 2007).

Research shows that in most cases both teachers and learners appear to have instrumental understanding of fractions, but do not really know why the procedures are used (Post, Harel, Behr, & Lesh, 1991). Students tend to develop conceptual schemes and information processing capacities to master fractions, decimals and percentage concepts individually but they also need to understand the commonalities between the different representations in their interaction with each other (Kieren, 1980). The educational aim however is for these students to have a balanced ability to follow a procedure with conceptual or relational understanding as the two facets interactively support a solid knowledge foundation (Zhou, 2011).

### Assessment and measurement

The rich theorising of and research into rational numbers provides the theoretical base for the assessment instrument, which therefore meets the requirement for measurement to define clearly what is to be tested (Wright & Stone, 1979, 1999). The next requirement is to outline the interrelationships between component parts of the construct; in the case of this study, the interrelationships between fraction, decimal and percentage representations. The third stage is the construction and selection of items that will operationalise the construct, keeping in mind its complexity, and which will provide the teacher with evidence of misconceptions that would need to be addressed in class. A final phase is the post hoc verification of the functioning of the test as a whole and of the individual items.

### Research design (participants, measures and models)

The primary study (Maseko, 2019) investigated the extent to which the 2015 cohort had mastered and retained their procedural and conceptual knowledge from their school level mathematics. This prior study reports on the level of relational understanding in the triad of concepts fractions-decimals-percentages of the first-year Foundation Phase student teachers entering the education programme. This article reports on the appropriateness of the instrument designed to test the students' levels of understanding and conceptual knowledge as they entered the teacher education programme.

The assessment tool was administered to the whole population of students that were admitted into the Foundation Phase teacher training programme ( $N = 117$ ). The test comprised 93 items that were designed to elicit prior knowledge at the beginning of the academic year.

The main research study comprised five conceptual categories that facilitated the analyses. The categories are understanding rational number concepts: definitions and conversions (14 items); manipulating symbols (operations) (17 items); comparing and sequencing rational numbers (15 items); alternate forms of rational number representation (35 items); as well as solving mathematical word problems with rational number elements (12 items). The items were drawn from selected projects, for example 'the rational number project' (Cramer, Behr, Post, & Lesh, 2009), and other such literature, and then adapted to post secondary school level.

The items were primarily informed by the conceptual categories above, and could be identified according to the following requirements:

- The items demanded a demonstration of procedural as well as conceptual understanding.
- The items included fraction, decimal and percentage representations.
- Items were generated with the specific purpose of evoking misconceptions.
- The items were comprehensive, covering most concepts and sub-concepts within the three representational systems – fractions, decimal fractions and percentages.
- The format of the test item types included multiple choice items, short answer, as well as extended response items.

The reason for such a comprehensive selection of items was that the lecturers needed to identify the many difficulties and misconceptions the students could bring into their first semester mathematics class. A range of difficulty that would include learners of current low proficiency, and high proficiency, was also required. Also, at the time of setting the items, the instructors were not sure from which categories the difficulties would emerge.

The Rasch model was applied in this study in order to either confirm or challenge the theoretical base, to check the validity of the instrument, and to measure the students' cognition of rational number concepts. The hypothesis was that the assessment tool would function according to measurement principles. The Rasch model provided information of where the item functioning and student responses were unexpected. Possible explanations could then be inferred, and presented, as well as provide some indications for the refinement of the test instrument.

### Methodology

There are other theories developed that can be used to validate and authenticate tests, such as Classical Test Theory (CTT) (Treagust, Chittleborough, & Mamiala, 2002). Rasch measurement theory (RMT) is generally used when measurement principles are considered, as the Rasch model adheres to measurement principles as conceptualised in the physical sciences (Rasch, 1960/1980). The application of the Rasch model is premised on the particular strength implicit in the model: that both item and person parameters are aligned on the same scale (Wei, Liu, & Jia, 2014). By

considering both the validity and reliability of the test items, and of the person responses, the Rasch model identifies aspects for further improvement as well as signs of biases (if any) (Smith, 2004; Bond & Fox, 2007, 2015; Long, Debba, & Bansilal, 2014; Bansilal, 2015).

## Ethical considerations

This study has been cleared by the University of Johannesburg Ethics Committee, with the ethical clearance number SEM 1 2018-021.

## Findings

The first analysis showed the test instrument to have a sound conceptual base and to be well targeted to the cohort, with a range of items, such that the students of current lower proficiency could answer a set of questions with relative ease, while students of high proficiency would experience some challenging items.

Table 1 shows summary statistics of the Rasch analysis. In this model, the item mean is set at zero, with items of greater and lesser difficulty calibrated against the mean. Person proficiency is then estimated against the item difficulty. The item standard deviation was 1.6302. The person mean location is estimated to be  $-0.4238$  logits, and the person standard deviation is 0.9686, which shows fairly good targeting and spread. The person separation index of 0.9114 shows that the assessment tool was able to differentiate well between students' proficiencies and that the power of fit was excellent, in essence a high reliability.

As observed in the person-item map Figure 1, a range of items from easy to difficult was achieved, and the test is well targeted.

Easier items are located at the lower end of the map (Item 65 and Item 66), while the difficult items are located at the higher end (Item 27 and Item 28). Similarly, learners of high proficiency are located higher on the map, 2.903 and 1.733, while learners of low proficiency are located at  $-2.159$  and  $-2.143$ . The mathematical structure of the Rasch model is such that where a person's proficiency location is aligned with an item difficulty location, an individual of that proficiency level has a 50% probability of answering an item of that difficulty level correctly (Rasch, 1960/1980). From the model one is able to predict how a student in a particular location will perform against an item: at, below or above their location on the scale.

**TABLE 1:** Summary statistics of fractions, decimals and percentages.

Statistic	Value
Item mean	0.0000
Item standard deviation	1.6000
Person mean	$-0.4238$
Person standard deviation	0.9686
Person separation index	0.9114
Power of analysis of fit	Excellent

## Individual item analysis

The individual items when constructed were initially reviewed by the lecturers. The application of the Rasch model provided empirical output calibrating a relative location and giving the probability that a person located at a certain proficiency location will get the item correct within the instrument.

Item 63 (43. Fraction form of 0.21) at position  $-0.646$  is shown on the category probability curve (Figure 2). Aligned with Item 63 are seven students (each represented by an  $\times$ , as shown on Figure 1). From their overall performance on the test as a whole, these students are estimated to have a 50% chance of answering Item 63 correctly. Each of the items shown by the category probability curves can be represented to show the item's unique characteristics in relation to the student cohort as a whole.

In Figure 2, depicting Item 63, the horizontal axis shows the student locations from  $-5$  to  $+5$ . The vertical axis indicates the probability of getting a correct response. The item difficulty is calibrated at  $-0.646$  (the dotted line shows the meeting point of the two curves). As stated previously, the seven students located at this point will have a 50% probability of answering the question correctly. Students located above  $-0.646$  will have a greater than 50% probability of answering this question correctly. Students located below  $-0.646$  will have a less than 50% chance of answering this question correctly. The light grey curve indicates the probability, according to the model, of a correct response. Inversely, the solid black curve shows the probability of getting an incorrect answer. Both curves plot either an increased or decreased probability of a correct response from a particular location of both a question item as well as a person responding to that item.

When an item is difficult or easy for the students, the curves show a shift of the meeting point away from the zero position (0) on the  $x$ -axis. Two items, Item 58, a relatively difficult item with an item location of about  $+3$  logits (see Figure 3), and Item 39, a relatively easy item, with an item location of about  $-3$ , are presented (see Figure 4).

Very few students are to the right of position  $+3$ , implying that it was only students located at  $+3$ , or higher, that had a greater than 50% probability of answering the item correctly.

Item 39 (Figure 4) had a 50% or greater probability of being answered correctly even by students with relatively low proficiency. All those to the right of location  $-3$  had a greater than 50% chance of providing the correct answer.

In this next discussion we compare two students, one located at  $+3$  and another located at  $-3$ , on Item 63 (Figure 5). The student located at  $+3$  has a 97% chance of answering this item correctly; however, the student located at  $-3$  has about a 4% of answering the item correctly. The model predicted this outcome, which is not to say that the low proficiency student





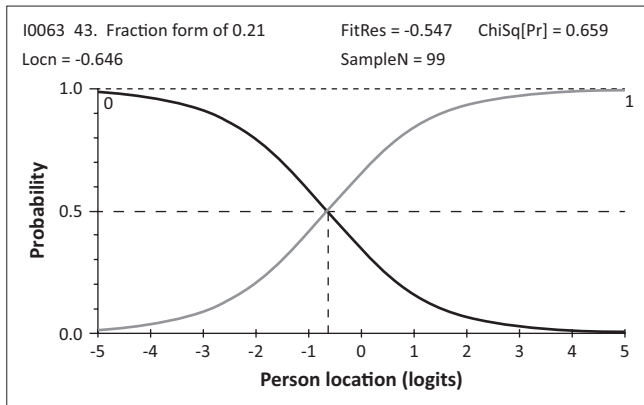
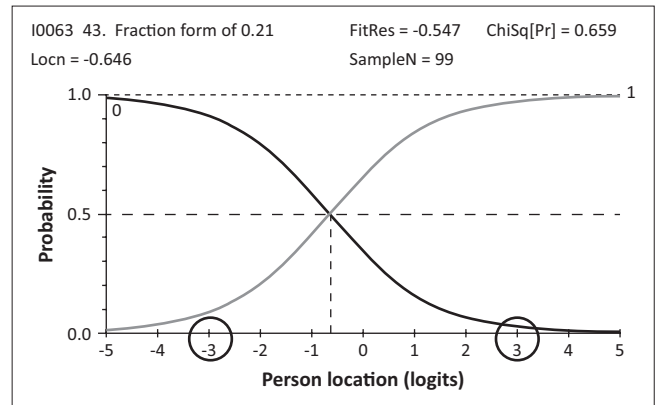


FIGURE 2: Item 63 – category probability curve.



Note: Relatively low proficiency (location -3) and highly proficiency (location 3).

FIGURE 5: Item 63 – category probability curve.

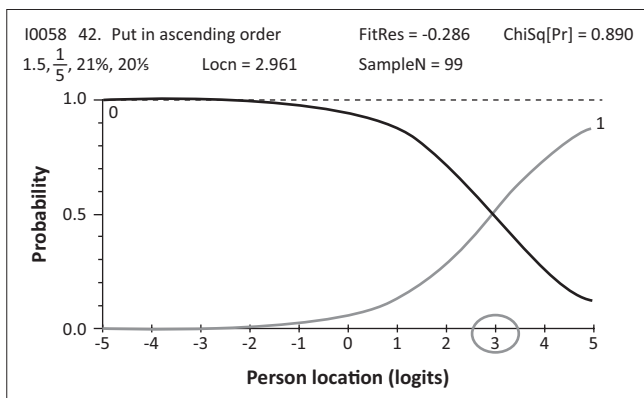


FIGURE 3: Item 58 – difficult category probability curve.

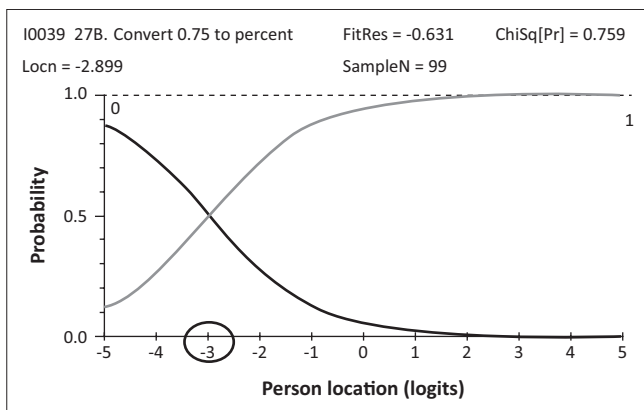


FIGURE 4: Item 39 – easy category probability curve.

could not answer a very difficult item correctly, but that this outcome was highly unlikely.

In summary, applying the Rasch model to a data set is essentially testing a hypothesis that invariant measurement has been achieved. Where there are anomalies, the researcher is required to investigate the threat to valid measurement. The model enables the researchers to identify the items that did not contribute to the information being sought or those items that were deemed faulty in some respect. Likewise, where students' responses to the question were unexpected the researchers were also alerted. The Rasch model is to some extent premised on the Guttman pattern, which postulates that in addition to some difficult questions, a person of

greater proficiency should answer all the items correctly that a person of lower proficiency answers correctly. Likewise, easier items should be answered correctly by low proficiency learners, and also by moderate proficiency and higher proficiency learners. While a strict Guttman pattern is not possible in practice, the principle is a good one (Dunne, Long, Craig, & Venter, 2012).

We briefly report on six students against four questions close enough to their locations to illustrate the relationship of person proficiency to item difficulty as seen through the Guttman pattern model.

The student of low proficiency (A, location  $-2.159$ ) struggled with the range of items that included the easiest of the items. The other student categorised as of low proficiency (B, location  $-2.143$ ), offered no response to these particular items. From the person-item map, we would expect students at these locations to have a 50% chance of answering correctly, meaning that if there were 100 students at that location approximately 50 could have answered the items correctly.

Of the two students in the moderate category, one of the students (C, location  $0.003$ ) did not attempt the easiest item (location  $-2.234$ ) (missing response), while the other student (D, location  $0.029$ ) answered this item far below his location correctly. The next two items which were above the two students' locations were either not answered or answered incorrectly.

The two students located in the high proficiency category are located at  $1.733$  logits (E) and  $2.903$  logits (F), more than a logit apart. We therefore deal with them separately. Student E answered the easiest item correctly and this was to be expected; however, the next easiest item was answered incorrectly. In theory the student should have had a greater than 50% correct response. The difficulty of the third item is aligned with the proficiency of Learner E. In theory Learner E has a 50% chance of answering Item 57 correctly. Item 58 has a greater difficulty by a large margin. One would expect the student to perhaps get this incorrect.

**TABLE 2:** Six students, low, moderate and high proficiency vs performance on four items.

Weakest to strongest student Easiest to difficult item	Low proficiency		Moderate proficiency		High proficiency	
	-2.159 (A)	-2.143 (B)	0.003 (C)	0.029 (D)	1.733 (E)	2.903 (F)
Item 79 (location -2.234)	0	MR	MR	1	1	1
Item 82 (location 0.0)	0	MR	0	1	0	1
Item 57 (location 1.733)	0	MR	MR	0	1	1
Item 58 (location 2.903) (Put in ascending order: $1.5, \frac{1}{5}, 21\%, 20 \frac{1}{5}$ )	0	MR	MR	0	0	0

MR, missing response.

Student F (location 2.903) answered three items correctly but was not able to answer Item 58 (location 2.903) correctly. According to the model the student had a 50% probability of answering this item correctly, as it is located at the same point on the scale. In the case of the most difficult item, Item 58, the requirement was to make decisions on converting the existing form before comparing and sorting the elements in ascending order. The cognitive demand required the students to connect their knowledge and make decisions in the process of working out the solution.

### Problematic items

It was noted in the first analysis that there were two items that did not function as expected. These two items were removed from this analysis, although for future testing they may be refined. One multiple choice item was removed due to an error. The second item, Question 8A, was revised as shown below and was reserved for the next cycle. Item 88 (Question 8A) was found to be a misfit as the grammatical representation of the mathematical idea is confusing. The original and possible revised versions are briefly discussed below.

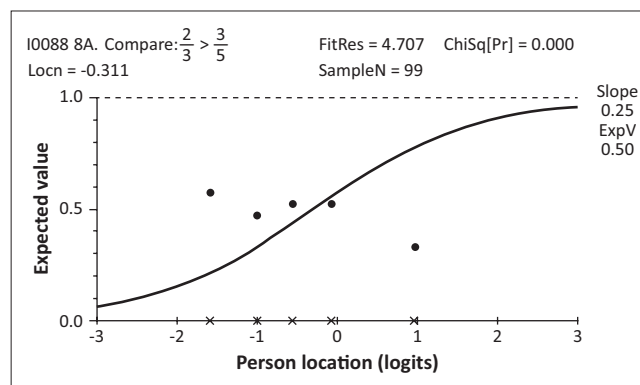
### Original question:

*Tell if the fraction on the left is less or greater than or equal to the fraction on the right. Use < or > or = for each case to make the statement true.*

$$8A. \frac{2}{3} \frac{3}{5} \quad 8B. \frac{6}{15} \frac{3}{5} \quad 8C. 1.5 \ 150\% \quad 8D. 5 \frac{3}{4} \ 5.75\%$$

The responses to item 8A produced the distribution displayed in Figure 6.

The black dots represent the means of the 5 class intervals into which the students were divided. The allocation to class intervals is decided by the researcher. The black dots representing students' mean responses did not follow the expected pattern according to the model. The expectation is that students of lower ability will be less likely to answer an item correctly than those of higher ability. The analysis revealed that learners of lower proficiency (four × marks left of 0 logits) on the test as a whole performed relatively higher than the students of higher proficiency (one × mark at about 1 logits). This anomaly was investigated, and it was found that the grammar and length of the instructions appeared to have interfered with the understanding of the question. For the next three items in Question 8 the instructions did not



**FIGURE 6:** Item 88 (8A) – item characteristic curve.

seem to mislead the students. When the instructions were revised and reduced to 'Use < or > or = for each case to make the statement true', the whole question seemed clearer.

### Local independence

A further check on the validity of the test required an investigation of local independence. In any test, one expects that each item would contribute some information to the test construct (Andrich & Kreiner, 2010). There may be cases of construct irrelevance, where items do not contribute to the construct, and may be testing other dimensions, or construct underrepresentation, where the construct is not fully represented (Messick, 1989). On the other hand, there may be cases where there is response dependency, where answering a second item correctly is dependent on answering the previous item correctly. Another threat to validity of the construct is where there are too many items targeting one aspect of the construct, for example five items asking for similar knowledge. In such a case the student who knows the concept is unduly advantaged, while a student who does not know the concept is unduly disadvantaged. High residual correlations between items can be resolved by forming a subset, essentially a super-item, where the two items contribute to the score (Andrich & Kreiner, 2010).

In this instrument analysis, we checked the residual correlations of the items and found high correlations, both positively correlated sets of items and negative correlations across some items. The implications of such a threat to local independence is that there are many items contributing the same information, as in a high positive correlation, and those with a negative correlation are 'pulling in the other direction'. A resolution of this threat is to remove the items that seem to test the same thing or create subtests of items that are highly

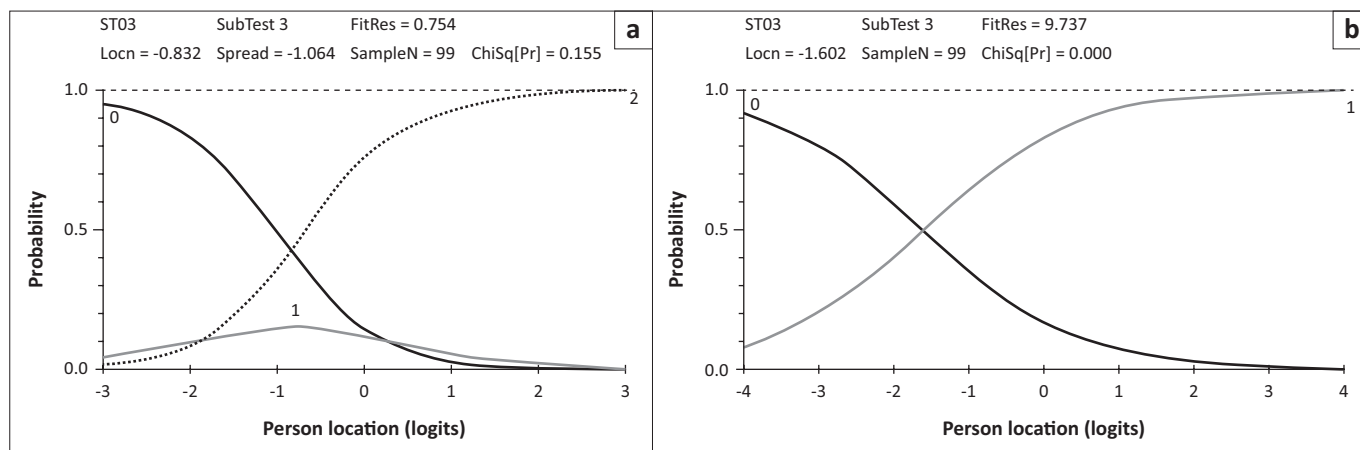


FIGURE 7: (a) Subtest 3 (question 6), (b) Subtest 3 (question 6) – re-scored.

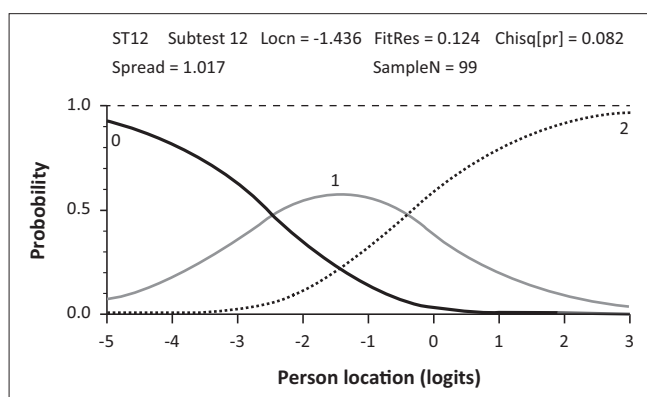


FIGURE 8: Subtest 12 (question 27A and 27B).

correlated, by investigating both the item context and the statistics it conveys.

In a second round, eight items were removed due to redundancy. In order to resolve response dependency, 18 subtests were created. These subtests were then checked for ordered or disordered thresholds. For illustrative purposes four sets of items are discussed.

Question 6: Item 6a 'Draw a representation of fraction  $\frac{2}{5}$ , and Item 6b 'Explain the meaning of the following fraction:  $\frac{2}{5}$ , were subsumed into a subtest. The subtest was structured in such a way that instead of having two items that were highly correlated, there was one partial credit item, for which the student could obtain a 0 for none correct, a 1 for one of the two questions correct, or a 2 for fully correct.

On investigating the subtest, Question 6 (combined a and b), which required students to both draw a representation and explain the meaning of  $\frac{2}{5}$ , the now partial credit item, it was observed that the common response was either none correct, or both correct. The middle category for which one mark awarded was almost redundant. The solution was to re-score the item as a dichotomous item and the resulting category probabilistic curve to show an improved scoring (Figure 7b).

Question 27 required the students to provide the fraction and percentage form for 0.75 as individual responses, but the correct answer depended on whether the student knew how to perform the conversions to both forms of fractions from the decimal form, that is, fraction and percentage form. This was the second set of items observed to be highly correlated and was subsumed into a subtest. For this subtest (see Figure 8) it was found that the category probability curves functioned appropriately. The three categories, 0, 1 and 2, corresponded to both incorrect, one correct and two correct. The group of students of middle proficiency were most likely to obtain 1 mark for being proficient in converting a decimal fraction to either a common fraction or a percentage, whereas the higher proficiency group obtained the full 2 marks, meaning that they were proficient in both conversions of the item.

The next subtest was created by subsuming four items into one set. The four sections of the question asked similar questions, which were to convert from an improper fraction to a mixed fraction. These four items – 11A =  $\frac{16}{5}$ ; 11B =  $\frac{18}{6}$ ; 11C =  $\frac{19}{4}$  and 11D =  $\frac{24}{5}$  – appear to be testing only one skill because the distribution showed that students either answered all four items correctly or answered none correctly. The resulting category probability curve is shown in Figure 9a. There may be a case here for rescoring, 0, 1 or 2 (see Figure 9b).

The final subtest was made up of four different question items, where the requirement was to order a combination of the fractions-decimals-percentages representations in ascending or descending order (See Figure 10).

Here it appeared that although these items were highly correlated, they increased in complexity. This subset functioned as expected in that the categories mark increase in proficiency with a clearer differentiated distribution of the curves (Figure 11).

As exhibited in the examples above, the investigation of specific subsets, from both a conceptual perspective and a statistical perspective, was conducted in order to ascertain which items could reasonably be subsumed into subtests.



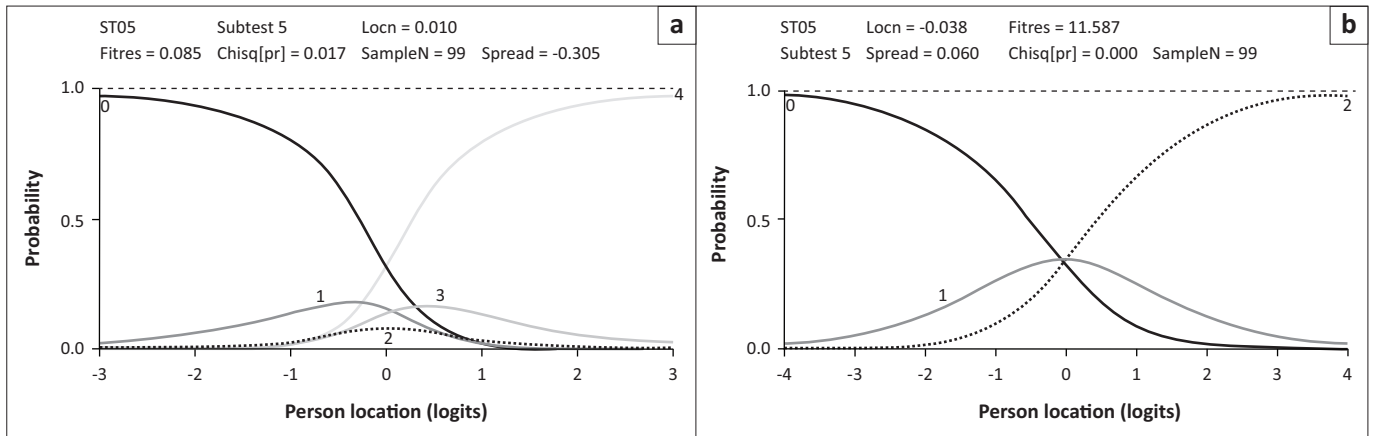


FIGURE 9: (a) Subtest 5 (question 11) – original score category probability curve, (b) Subtest 5 (question 11) – re-scored category probability curve.

<b>39. Order from least to greatest:</b> A: 13.6%, B: 20.25, C: 316%, D: 30%	<b>40. Order from greatest to least:</b> A: 0.9, B: %, C: 81%, D: 1%
<b>41. Put in descending order:</b> A: 33.3%, B: 0.33, C: 0.30, D: 33%	<b>42. Put in ascending order:</b> A: 1.5, B: %, C: 21%, D: 20%

FIGURE 10: Final subtest.

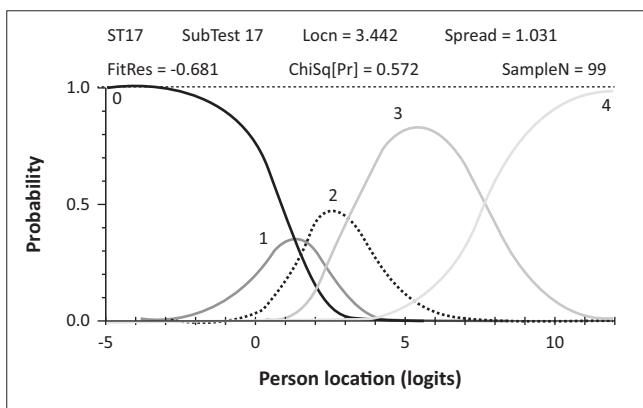


FIGURE 11: Subtest 17 (questions 39–42) – re-scored category probability curve.

The subtests that functioned as expected were retained, but for those whose categories were for some conceptual reason not functioning according to measurement principles, the rescoring of the subtests items was implemented. The process reported in this article works together with a qualitative investigation that was done in the main study, and also formed part of improving the functioning of the instrument (Dunne et al., 2012; Maseko, 2019).

The outcome, after this final analysis, was a test with 50 items, including both dichotomous and polytomous items, 22 of which were multiple choice format and 28 constructed response format. Figure 12 appears more compact in the distribution of both the test item difficulty locations and students' proficiency levels. The easiest of the items (ST031) is by far the easiest and there is some distance from the next easiest items by almost 2.5 logits points. There were two items that were at difficulty levels where no one had a 50%, or greater, chance of answering correctly.

Table 3 shows the refined test person mean; a mean of  $-0.4172$  in the initial analysis, moved closer to zero,  $0.099$ , implying

that by resolving some of the test issues, the targeting of the test to the learners was found to be better. The item standard deviation in the initial analysis was rather large at 1.6302, but after the refinement of the test it moved closer to 1, at 1.4933. The person standard deviation after refinement was somewhat smaller, implying that the range of proficiency was narrower.

## Conclusion and implications

As stated in the introduction, this article forms part of a larger study into the student understanding of rational number, fractions, decimals and percent. The purpose of the investigation was to gather information about the cohort entering the Foundation Phase teacher development on working with rational numbers, especially fractions-decimals-percentage. This article reported on how the instrument was functioning to assess their knowledge level of work done at school. The assessment tool covered understanding rational number concepts, manipulating symbols (operations), comparing and sequencing rational numbers, alternate forms of rational number representation, as well as solving mathematical word problems with rational number elements. It is clear that the number of items does not impact the quality of the test. Beyond a certain amount, some of the items might be redundant. One has to check if the test instrument as a whole is fit for purpose. Beyond the total score obtained by each student in the test, the Rasch model indicates a position on a unidimensional scale where the student's proficiency level is differentiated. The power and usefulness of the Rasch model is that it supports the professional judgement of the subject expert in making decisions about the validity of items (Smith & Smith, 2004).

The Rasch model was applied in this study in order to confirm or challenge the theoretical base, to check the validity of the instrument and to quantify the students' cognition of rational number concepts.

The application of the Rasch measurement model enabled checking whether the test content was consistent with the construct under investigation, and supported expectations of a sharper understanding of these students in terms of



**TABLE 3:** A comparison of the initial and final analyses.

Initial analysis		Final analysis	
Item mean	0.0000	Item mean	0.0000
Item standard deviation	1.6302	Item standard deviation	1.4933
Person mean	-0.4172	Person mean	0.0990
Person standard deviation	0.9442	Person standard deviation	0.7580
Person separation index	0.9088	Person separation index	0.8606
Power of analysis of fit	Excellent	Power of analysis of fit	Excellent

proficiency level within a set of items in the test. The outcome showed the data to fit the model, the person separation index was high, and the target was appropriate, thereby confirming the theoretical work that supported the design of the test.

## Acknowledgements

This work is based on a doctoral study for which the University of Johannesburg has provided financial support.

The opinions, findings, conclusions and recommendations expressed in this manuscript are those of the authors and do not necessarily reflect the views of the University of Johannesburg.

## Competing interests

The authors declare that they have no personal relationships that may have inappropriately influenced the writing of this manuscript.

## Authors' contributions

This study is based on the submitted doctoral research of J.M. The conceptualisation of this manuscript was done by all three authors. K.L. provided input on the literature and methodology. C.L. assisted with the Rasch analysis. The final product received input from all three authors.

## Funding information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## Data availability statement

Data sharing is not applicable to this article as no new data were created or analysed in this study.

## Disclaimer

The opinions, findings, conclusions and recommendations expressed in this manuscript are those of the authors and do not necessarily reflect the views of the University of Johannesburg.

## References

- Andrich, D., & Kreiner, S. (2010). Quantifying response dependence between two dichotomous items using the Rasch model. *Applied Psychological Measurement, 34*(3), 181–192. <https://doi.org/10.1177/0146621609360202>
- Andrich, D., Sheridan, B.S., & Luo, G. (2013). *RUMM2030: An MS Windows program for the analysis of data according to Rasch Unidimensional Models for Measurement*. Perth: RUMM Laboratory. Retrieved from <http://www.rummlab.com.au>

- Bansilal, S. (2015). A Rasch analysis of a Grade 12 test written by mathematics teachers. *South African Journal of Science, 111*(5–6), 1–9. <https://doi.org/10.17159/sajs.2015/20140098>
- Bond, T.G., & Fox, C.M. (2007). *Applying the Rasch model: Fundamental measurement in the human science* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bond, T.G., & Fox, C.M. (2015). *Applying the Rasch model: Fundamental measurement in the Human Sciences* (3rd ed.). New York, NY: Routledge. <https://doi.org/10.4324/9781315814698>
- Charalambous, C.Y., & Pitta-Pantazi, D. (2005). Revisiting a theoretical model on fractions: Implications for teaching and research. In H.L. Chick & J.L. Vincent (Eds.), *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 233–240). Melbourne: PME. Retrieved from <https://www.emis.de/proceedings/PME29/PME29RRPapers/PME29Vol2CharalambousEtAl.pdf>
- 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>
- Cramer, K., Behr, M., Post, T., & Lesh, R. (2009). *Rational number project: Initial fraction ideas*. Retrieved from <http://wayback.archive-it.org/org-121/20190122152926/http://www.cehd.umn.edu/ci/rationalnumberproject/rnp1-09.html>
- Dunne, T., Craig, T., & Long, C. (2012). Meeting the requirements of both classroom-based and systemic assessment of mathematics proficiency: The potential of Rasch measurement theory. *Pythagoras, 33*(3), Art. #19. <https://doi.org/10.4102/pythagoras.v33i3.19>
- Gabriel, F.C., Szucs, D., & Content, A. (2013). The development of the mental representations of the magnitude of fractions. *PLoS One, 8*(11), e80016. <https://doi.org/10.1371/journal.pone.0080016>
- Harvey, R. (2011). Challenging and extending a student teacher's concepts of fractions using an elastic strip. In J. Clark, B. Kissane, J. Mousley, T. Spencer, & M. Thornton (Eds.), *Proceedings of the 34th Conference of the Mathematics Education Research Group of Australasia* (pp. 333–339). Adelaide: MERGA. Retrieved from [https://merga.net.au/Public/Public/Publications/Annual\\_Conference\\_Proceedings/2011\\_MERGA\\_CP.aspx](https://merga.net.au/Public/Public/Publications/Annual_Conference_Proceedings/2011_MERGA_CP.aspx)
- Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge: The case of mathematics. In J. Hiebert (Ed.), *Conceptual and procedural knowledge in mathematics: An introductory analysis* (Vol. 2, pp. 1–27). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Huang, T.-W., Liu, S.-T., & Lin, C.-Y. (2009). Preservice teachers' mathematical knowledge of fractions. *Research in Higher Education Journal, 5*(1), 1–8. Retrieved from <https://www.aabri.com/manuscripts/09253.pdf>
- Kieren, T.E. (1980). *Five faces of mathematical knowledge building*. Edmonton: Department of Secondary Education, University of Alberta.
- Long, C., Debba, R., & Bansilal, S. (2014). An investigation of Mathematical Literacy assessment supported by an application of Rasch measurement. *Pythagoras, 35*(1), Art #235. <https://doi.org/10.4102/pythagoras.v35i1.235>
- Maseko, J. (2019). *Exploring the most common misconceptions and the associated errors that student teachers at foundation phase display when studying fractions for teaching*. Unpublished doctoral dissertation, University of Johannesburg, Johannesburg, South Africa. Retrieved from <http://hdl.handle.net/10210/398134>
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5–11. <https://doi.org/10.3102/0013189X018002005>
- Pantziara, M., & Philippou, G. (2012). Levels of students' 'conception' of fractions. *Educational Studies in Mathematics, 79*(1), 61–83. <https://doi.org/10.1007/s10649-011-9338-x>
- Parker, M., & Leinhardt, G. (1995). Percent: A privileged proportion. *Review of Educational Research, 65*(4), 421–481. <https://doi.org/10.3102/00346543065004421>
- Pesek (Simoneaux), D.P., Gray, E.D., & Golding, T.L. (1997). *Rational numbers in content and methods courses for teacher preparation*. Washington, DC: Educational Resource Center (ERIC). Retrieved from [https://ia802903.us.archive.org/32/items/ERIC\\_ED446934/ERIC\\_ED446934.pdf](https://ia802903.us.archive.org/32/items/ERIC_ED446934/ERIC_ED446934.pdf)
- Post, T.R., Harel, G., Behr, M., & Lesh, R. (1991). Intermediate teachers' knowledge of rational number concepts. In E. Fennema, T.P. Carpenter, & S.J. Lamon (Eds.), *Integrating research on teaching and learning mathematics* (pp. 177–198). New York, NY: State University of New York Press.
- Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests* (Expanded edition with foreword and afterword by B.D. Wright). Chicago, IL: University of Chicago Press.
- Ryan, J., & Williams, J. (2007). *Children's mathematics 4-15: Learning from errors and misconceptions*. Berkshire: McGraw-Hill Education.
- Sangwin, C.J. (2007). Assessing elementary algebra with STACK. *International Journal of Mathematical Education in Science and Technology, 38*(8), 987–1002. <https://doi.org/10.1080/00207390601002906>
- Shalem, Y., Sapire, I., & Sorto, M.A. (2014). Teachers' explanations of learners' errors in standardised mathematics assessments. *Pythagoras, 35*(1), Art #235. <https://doi.org/10.4102/pythagoras.v35i1.254>
- Smith, E., & Smith, R. (2004). *Introduction to Rasch Measurement: Theory, models and applications*. Maple Grove, MN: JAM Press.
- Smith, R.M. (2004). Detecting item bias with the Rasch model. *Journal of Applied Measurement, 5*(4), 430–449.
- Stacey, K., Helme, S., Steinle, V., Baturo, A., Irwin, K., & Bana, J. (2001). Preservice teachers' knowledge of difficulties in decimal numeration. *Journal of Mathematics Teacher Education, 4*(3), 205–225. <https://doi.org/10.1023/A:1011463205491>
- Stiggins, R., & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory into Practice, 44*(1), 11–18. [https://doi.org/10.1207/s15430421tip4401\\_3](https://doi.org/10.1207/s15430421tip4401_3)

- Treagust, D.F., Chittleborough, G., & Mamiala, T.L. (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education, 24*(4), 357–368. <https://doi.org/10.1080/09500690110066485>
- Vamvakoussi, X., & Vosniadou, S. (2007). How many numbers are there in a rational number interval? Constraints, synthetic models and the effect of the number line. In S. Vosniadou, A. Baltas, & X. Vamvakoussi (Eds.), *Reframing the conceptual change approach in learning and instruction* (pp. 265–282). Oxford: Elsevier.
- Vamvakoussi, X., & Vosniadou, S. (2010). How many decimals are there between two fractions? Aspects of secondary school students' understanding of rational numbers and their notation. *Cognition and Instruction, 28*(2), 181–209. <https://doi.org/10.1080/07370001003676603>
- Venkat, H., & Spaul, N. (2015). What do we know about primary teachers' mathematical content knowledge in South Africa? An analysis of SACMEQ 2007. *International Journal of Educational Development, 41*, 121–130. <https://doi.org/10.1016/j.ijedudev.2015.02.002>
- Wei, S., Liu, X., & Jia, Y. (2014). Using Rasch measurement to validate the instrument of Students' Understanding of Models in Science (SUMS). *International Journal of Science and Mathematics Education, 12*(5), 1067–1082. <https://doi.org/10.1007/s10763-013-9459-z>
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Wright, B.D., & Stone, M.H. (1979). The measurement model. In B.D. Wright & M.H. Stone (Eds.), *Best test design* (pp. 1–17). Chicago, IL: MESA Press.
- Wright, B.D., & Stone, M.H. (1999). *Measurement essentials*. Wilmington, DE: Wide Range, Inc.
- Zhou, Z. (2011). The clinical interview in mathematics assessment and intervention: The case of fractions. In M.A. Bray & T.J. Kehle (Eds.), *The Oxford handbook of school psychology* (pp. 351–368). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195369809.013.0131>