



# Item difficulty analysis of a high-stakes mathematics examination using Rasch analysis

## Authors:

Mark Jacobs<sup>1</sup>  
Duncan Mhakure<sup>2</sup>  
Richard L. Fray<sup>3</sup>  
Lorna Holtman<sup>4</sup>  
Cyril Julie<sup>5</sup>

## Affiliations:

<sup>1</sup>Electrical Engineering Department, Faculty of Engineering, Cape Peninsula University of Technology, South Africa

<sup>2</sup>Numeracy Centre, University of Cape Town, South Africa

<sup>3</sup>Department of Mathematics and Applied Mathematics, University of the Western Cape, South Africa

<sup>4</sup>Postgraduate Studies, University of the Western Cape, South Africa

<sup>5</sup>School of Science and Mathematics Education, University of the Western Cape, South Africa

## Correspondence to:

Mark Jacobs

## Email:

[jacobsms@cput.ac.za](mailto:jacobsms@cput.ac.za)

## Postal address:

PO Box 1906, Bellville 7535, South Africa

## Dates:

Received: 21 Feb. 2013

Accepted: 23 Mar. 2014

Published: 21 May 2014

## How to cite this article:

Jacobs, M., Mhakure D., Fray, R.L., Holtman, L., & Julie, C. (2014). Item difficulty analysis of a high-stakes mathematics examination using Rasch analysis. *Pythagoras*, 35(1), Art. #220, 7 pages. <http://dx.doi.org/10.4102/pythagoras.v35i1.220>

## Read online:



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

The National Senior Certificate examination is the most important school examination in South Africa. Analysis of learners' performance in Mathematics in this examination is normally carried out and presented in terms of the percentage of learners who succeeded in the different bands of achievement. In some cases item difficulties are presented – item refers to the subsection of each examination question. Very little attention is paid to other diagnostic statistics, such as the discrimination indices and item difficulties taking into consideration partial scores examinees achieve on items. In this article we report on a study that, in addition to the usual item difficulties, includes a discrimination index of item difficulties taking into account partial scores examinees achieved. The items, considered individually, are analysed in relation to the other items on the test. The focus is on the topic sequences and series and the data were obtained from a stratified sample of the marked scripts of the candidates who wrote the National Senior Certificate examination in Mathematics in November 2010. Rasch procedures were used for the analysis. The findings indicate that learners perform differently on subsections of topics, herein referred to as items, and that focusing on scores for full topics potentially mask these differences. Mathematical explanations are attempted to account for difficulties learners exhibit in these subsections, using a hierarchy of scale. The findings and our analysis indicate that a form of measurement-driven testing could have beneficial results for teaching. Also, for some items the difficulty obtained from the work of examinees runs counter to the commonly perceived wisdom that an examination ought to be structured in such a way that the less difficult items are at the start of a topic. An explanatory device anchored around the construct of 'familiarity with problem types through repeated productive practice' is used to account for the manifested hierarchy of difficulty of the items.

## Introduction

In most countries schooling culminates with learners having to write examinations in various subjects in order to obtain a certificate to matriculate. In South Africa the matriculation or the final National Senior Certificate (NSC) examination is a high-stakes examination and the outcome is a public event in which results are announced by the Minister of Basic Education and published in national newspapers. The NSC results can offer access to higher certificate programmes, diploma courses and the much sought-after degree courses at higher education institutions in the country. Mathematics or Mathematical Literacy is taken by all learners as a qualifying subject in the NSC examination. Mathematics is seen as the gatekeeper subject to access many degree programmes as Mathematical Literacy has only been differentially accepted by higher education institutions and not at all in most cases to degree programmes requiring Mathematics. A point score system is used at most universities in South Africa and the Mathematics score is usually weighted higher (it is doubled) in the score.

In this article we are interested in the extent to which learners experience difficulty with the topics in such high-stakes examinations. In particular, we analyse learner performance in sequences, series and the accompanying finance, growth and decay applications as evidenced by their performance in the high-stakes NCS Mathematics examination.

## Examinations

It is well known that there are different kinds of examinations serving different purposes. For example, there are criterion-referenced, norm-referenced, formative and summative examinations. In terms of the procedures employed for the construction, the mechanisms used for assessing the responses of examinees, the methods of quality control of assessment of examinees' responses (known as moderation) and the major beneficiaries of the examination, a simple classification

Copyright: © 2014. The Authors. Licensee: AOSIS OpenJournals. This work is licensed under the Creative Commons Attribution License.



along the lines of external or internal and examinee or system as the major beneficiary can be made. Table 1 presents such a classification for different examinations operative in the South African schooling system.

In Table 1, internal and external refer to whether the teacher or a group of teachers who teach Mathematics at the institution have the responsibility for the particular procedure of the examination process or not. If the responsible teacher is totally responsible but there is some form of external oversight as is the case, for example, with the teacher-constructed end-of-year (grade) examination where oversight is exercised by curriculum advisors, then this is indicated by '±'. With respect to beneficiaries, the examinee is the major beneficiary if the level of achievement has consequences for the examinee in terms of what worth it is to them. So, for example, attaining a level 4 (50% – 59%) pass in Mathematics in the NSC examination will allow access to a bachelor's degree if the examinee also performs at pre-determined levels in some other subjects. Where the system is the beneficiary, an examination is essentially a mechanism to supply bureaucrats and politicians with information regarding the effectiveness of the entire system and has no direct consequence for examinees in terms of how it will impact their immediate progression to, say, a next grade or obtaining a level of achievement that will allow access to various forms of further studies. As is clear from Table 1 the NSC examination is one of a few high-stakes examinations which is entirely independent of those teaching the subject at hand on all the processes involved in the examination process. In this sense it is on par with examinations for entry into high-status professions such as the board examination to become a chartered accountant.

## Number patterns and sequences in the curriculum

In mathematics, curriculum developers both in national and international arenas have identified the teaching and learning of number patterns as one of the main aims of mathematics. Vogel (2005, p. 445) argues that 'the analysis of number patterns and the description of their regularities and properties is one of the aims of mathematics'. According to Devlin (1996) the question of 'What constitutes mathematics?' has been the subject of much discussion in the 19th century. However, the definition of mathematics accepted by mathematicians only emerged three decades ago:

**TABLE 1:** Classification for different examinations operative in the South African schooling system.

Type of procedure	Nature of procedure	End-of-year (grade) examination	Annual national assessments	Systemic test	National Senior Certificate
Examination construction procedure	Internal	+			
	External		+	+	+
Assessment (mark allocation) procedure	Internal	+	+		
	External			+	+
Moderation procedure	Internal	±	±		
	External	±	±		+
Major beneficiary	Examinee	+			+
	System		+	+	

'mathematics is the science of patterns' (Devlin, 1996, p. 3). Given this definition, there is no doubt of the importance of the teaching and learning of number patterns within the South African national curriculum.

In this study we selected to focus the analysis on the topics of sequences and series with their organic relationship with number patterns and the closely related applications to finance, growth and decay. In the Curriculum and Assessment Policy Statement, the topics are labelled 'Number patterns, sequences, series' and 'Finance, growth and decay' (Department of Basic Education, 2012, p. 9). These two topics account for about 27% of the marks allocated in the first paper of the NSC examination for Mathematics.

The learning of patterns, sequences and series does not start in Grade 12. It commences in the Foundation Phase where it is stated: 'In this phase, learners work with both number patterns (e.g. skip counting) and geometric patterns (e.g. pictures)' (Department of Basic Education, 2011, p. 9). The topic continues through the Intermediate Phase and Senior Phase. In Grades 10–12, at the Further Education and Training phase, the exploration of number patterns is expanded and emphasis shifts to include goals that will allow learners to be able to:

- investigate number patterns leading to those where there is a constant difference between consecutive terms, and the general term is therefore linear
- investigate number patterns leading to those where there is a constant second difference between consecutive terms, and the general term is therefore quadratic
- identify and solve problems involving number patterns that lead to arithmetic and geometric sequences and series, including infinite geometric series (Department of Basic Education, 2012, p. 12).

In sum, the teaching and learning of number patterns plays a significant role in mathematics in schools. Firstly, Mason (1996) argues that learning number patterns develops learners' abilities in expressing generality, which is key to studying algebra and abstract mathematics. Herbert and Brown (1997, p. 126) express the same sentiments as Mason: 'From generalizing the pattern, students understand the power of algebraic thinking'. Secondly, recent lines of research (Lee, Bull, Ng, Pe & Ho, 2011) show that generating additional numbers in sequences enhances learners' arithmetic knowledge of computation and that proficiency in number patterns greatly helps in achieving good performance in algebra. Research has also shown that algebra has foundations in number patterns and that most of the challenges experienced by learners when learning algebra are as a result of lack of proper foregrounding in number concepts (Carragher & Schliemann, 2007). Lastly, if number patterns are taught using socio-cultural contexts then learners will develop notions that there should be patterns or norms in our daily life activities, and that the disruptions to these patterns should signify new challenges whose solutions need to be found.



Not only is the topic of patterns, sequences and series of importance for school mathematics from a historical perspective, it also underpinned Newton's work and its extension by Lagrange on the difference formula to determine polynomial functions fitting given sets of numbers (see e.g. Cuoco, 2005). In addition many non-polynomial functions, such trigonometric functions can be represented as power series, as for example, a Taylor series expansion of  $\sin x$ :

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots = \sum_{n=0}^{\infty} (-1)^{n+1} \frac{x^{2n+1}}{(2n+1)!}$$

The topic thus stretches its tentacles deep into post-school mathematics.

### The focus of this research

The issue pursued in this article therefore is the difficulty learners find with sequences, series and the accompanying finance, growth and decay topics as evidenced by their performance in the high-stakes NCS Mathematics examination. Analysis of learner performance in this examination is normally done at a global level. At this level feedback is provided of performance in an entire question, which normally deals with a topic as given in the prescribed curriculum document. The procedure followed to determine the difficulty of a question is of a classical test theory nature in which the emphasis is on the percentage of examinees who answered a question correctly. No consideration is given to subsections of a question and scant attention is given to partial scores a candidate obtains.

The research reported here is underpinned by the assumption that insightful information about learners' proficiency in high-stakes mathematics examinations can be gained by analysing test scores, taking into account scores obtained in subsections of questions on a topic. It would therefore be a form of measurement-driven testing leading to teaching that is directed towards achievement in these kinds of examinations, the antithesis of psychological and curriculum-driven testing.

### Procedure and methods

The study reported here is part of a larger study in process dealing with learners' ways of working in high-stakes school examinations in Mathematics. This larger study focuses on a 12% random sample of the scripts of examinees in the Western Cape who wrote the 2010 NCS Mathematics examination. The sample was stratified according to former (pre-1994) Cape Education Department schools, Department of Education and Training and House of Representatives schools, new schools established post-1994 and independent schools as well as the eight school districts in the Western Cape. For logistical reasons of not having to go through all the scripts in the province to obtain individual scripts, scripts from an entire school (examination centre) were selected. This rendered 1959 scripts. From this collection of scripts a similar random sample of 1122 scripts from the school were selected for the study reported here.

### Item analysis

Analysing items is an investigation of 'the performance of items considered individually either in relation to some external criterion or in relation to the remaining items on the test' (Thompson & Levitov, 1985, p. 163). For this article the focus is on the latter. The statistics given are the item difficulties, the Rasch item difficulties and the discrimination coefficients. These are the usual diagnostic statistics, which are calculated to determine how a particular test or subset of a test is functioning in order to ascertain the appropriateness of the test for a particular cohort. The item difficulty is the percentage of examinees who were awarded full marks for an item. The Rasch measure is the item difficulty obtained by applying a Rasch partial credit analysis to determine how both items and examinees are ranked in order of the difficulty level of the items. In this article we are not discussing the technicalities of Rasch modelling. There exists a rich corpus of literature pertaining to this and Dunne, Long, Craig and Venter (2012), for example, provide a comprehensive description of such technicalities. See Dunne et al also for a fuller description of the Rasch measurement theory where it is used to support both classroom-based and systemic assessment of mathematics proficiency. In particular, they show how systemic assessment using this form of analysis can lead to more informed teaching in the classroom.

For our purposes we use the Rasch measurement to analyse within items, that is, subsections of items, taken from sequences and series. For this paper Rasch analysis was done using the WINSTEPS Version 3.65.0 software. As alluded to above 'the Rasch measure for items is the item difficulty' (Linacre, 2008, p. 362). We report the measure for the items. 'The unit of measure used by Rasch for calibrating' (Linacre, 2008, p. 362) items is obtained is by logs-odds scaling ( $\ln \text{odds} = \ln \frac{P}{1-P}$ , where P = percentage answered correctly)

and reported in logits. The discrimination coefficient is the point-polyserial correlation between the examinee's score for an item and their total score excluding the score for the item that is correlated. It is the Pearson product-moment correlation and gives an indication of whether higher scores correspond to higher achievement for the test or a coherent subsection thereof. As a rule of thumb, high discrimination coefficients indicate that such items indicate higher performing examinees. In essence the discrimination coefficient is a measure that identifies a good item: an item that will discriminate between high and low scorers.

The advantage of using the discrimination coefficient instead of other forms of discriminative measures (such as the discriminative index) is that the former can be used with less than the full group of examinees (Backhoff, Larrazolo & Rosas, 2000). This is applicable in the case of this paper as a selection of the scripts of the total number of examinees who sat for the Mathematics 2010 Paper 1 was used. Diagnostic statistics were computed for each item, by which we mean the subsection of each question. The statistics reported are

the difficulty level, the discrimination coefficient and the Rasch item difficulty. These are the usual diagnostic statistics reported for large scale international tests such as, for example, TIMSS (Mullis, Martin, Gonzales & Chrostowski, 2004).

## Findings

Table 2 presents the diagnostic statistics for the sample of examinees' scripts selected for this study.

Items were identified as 'not attempted' when there was no indication on the script that the examinee had answered the question. The examinee either only wrote down the item number but the script indicated no further work with it or the item did not appear at all. This is distinguished from a '0 (zero)' mark where there were indications that the question was attempted or answered and a mark of '0' was awarded for the item by the markers. By indicating the 'not attempted', we are following the customary format of reporting diagnostic statistics as used in large-scale testing such as is done in the TIMSS reports (Mullis et al., 2004). Reporting the 'not attempted' gives a sense of completeness.

Table 2 shows the item difficulty order for both the item difficulty and the Rasch difficulty measure. The order slightly differs for the two measures of difficulty. This is to be expected since for the item difficulty only the examinees who scored full marks were taken into account to determine the index. The partial credit model used in calculating the Rasch measure takes into account all marks for an item. For example, Item 2.2.1 had an allocation of 3 marks and an examinee could be awarded a mark of 0, 1, 2 or 3. The Rasch procedure used the mark the examinee was assigned and the item difficulty index calculation does not take partial marks into reckoning.

The person-item map (Figure 1) represents the Rasch measures in order of difficulty. The right-hand side is a

hierarchical ordering of the difficulty of the items with the most difficult item at the top and the easiest item at the bottom. The number of examinees who had success at a particular level of difficulty is on the left. In essence the right-hand side gives an indication of the number of examinees who had at least a 50% chance of succeeding on items of similar difficulty. As is indicated in the figure, the '#' indicates 11 examinees and '.' indicates from 1 to 10 examinees. So, for example, there were, at most, 142 learners who had at most a 50% chance of succeeding on items of a similar kind to Item 2.3.2. These same learners had a less than 50% chance to succeed on items of the kind above 2.3.2 (Items 2.2.1, 3.1, 3.2, 7.1, 7.2.1, 7.2.2 and 7.2.3). They had a more than 50% chance

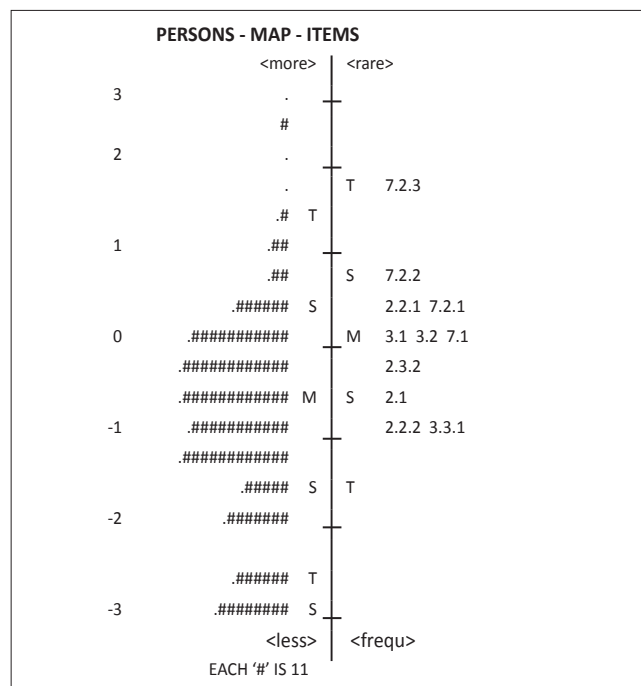


FIGURE 1: Person-item-map.

TABLE 2: Diagnostic statistics for the selected scripts and items.

Question	Item difficulty	Rasch measure	Discrimination coefficient	Not attempted
2.1 Evaluate: $\sum_{n=1}^{20} 3^{n-1}$	24	-0.64	0.66	6%
2.2.1 The following sequence forms a convergent geometric sequence: $5x; x^2; \frac{x^3}{5}; \dots$ Determine the possible values of $x$ .	8	0.45	0.56	18%
2.2.2 If $x = 2$ , calculate $S_{\infty}$ .	42	-1.04	0.58	11%
2.3.1 The following arithmetic sequence is given: 20; 23; 26; 29; ...; 101. How many terms are there in this sequence?	54	-1.11	0.48	3%
2.3.2 The even numbers are removed from the sequence. Calculate the sum of the terms of the remaining sequence.	22	-0.49	0.64	10%
3.1 The sequence $4; 9; x; 37; \dots$ is a quadratic sequence. Calculate $x$ .	22	0.03	0.65	19%
3.2 Hence, or otherwise, determine the $n^{\text{th}}$ term of the sequence.	17	0.02	0.68	29%
7.1 At what annual percentage interest rate, compounded quarterly, should a lump sum be invested in order for it to double in 6 years?	14	0.06	0.70	27%
7.2 Timothy buys furniture to the value of R10 000. He borrows the money on 01 February 2010 from a financial institution that charges interest at a rate of 9,5% p.a. compounded monthly. Timothy agrees to pay monthly instalments of R450. The agreement of the loan allows Timothy to start paying these equal monthly instalments from 01 August 2010.)	11	0.48	0.54	9%
7.2.1 Calculate the total amount owing to the financial institution on 01 July 2010.				
7.2.2 How many months will it take Timothy to pay back the loan?	5	0.63	0.68	19%
7.2.3 What is the balance of the loan immediately after Timothy has made the 25 <sup>th</sup> payment?	2	1.62	0.57	40%



of succeeding on items of the kind below 2.3.2 (Items 2.1, 2.2.2 and 2.3.1).

Julie (2012) developed an organisational scheme to cluster items into four zones of difficulty (see Table 3). This clustering is around the mean Rasch measure and the standard deviation. These zones and the items appearing in them, based on the item map, are presented in Table 4.

The moderately high and moderately low zones are volatile and it can be expected that items will, over time, transition between the zones (Julie, 2012). With this in mind, the hierarchy of the items as per the classical item difficulty index and the Rasch difficulty measure are the same.

## Discussion

For this quantitative study the only data we had were the scores as reflected on the scripts of the learners. We saw a particular pattern in terms of difficulty according to the model postulated by Julie (2012). As an organiser we looked at the nature of the question and tried to generate tentative explanations for the placement of the item in those particular zones of difficulty. The explanations are anchored around three constructs: familiarity with problem types through repeated productive practising, procedural complexity and interpretive complexity. Familiarity with solution of problem types is akin to the East Asian perspective that repeated practice is a necessary condition for the development of 'cleverness and creativity' (Julie et al., 2010, p. 366). Productive practising (Selter, 1996) is the notion that repeated practice is not just senseless drill for automated responses to cue-based questions. It includes drilling for mastery as well as activities for developing learners' mathematical thinking skills of the constructs that are being drilled. It is thus different from the kind of drill in which, for preparation for the NSC Mathematics examination, learners only work through previous examination papers. Watson and Mason (1998, p. 20), for example, pose the question 'is it always true that operating the same way on both sides of an equation produces an equivalent equation?' as a task to develop the skills of explaining,

justifying, verifying and refuting when learners are engaged with developing procedural fluency with the techniques for solving polynomial equations. Procedural complexity is 'the number of decisions and operations and sub problems' (Watson & De Geest, 2012, p. 223). Interpretive complexity refers to the learners' comprehension of a problem statement. It is not restricted to 'word problems' or contextdriven tasks. It encompasses instruction-type statements such as 'show that  $expression A = expression B$ ', where  $expression A$  and  $expression B$  are mathematically equivalent. This kind of instruction-type statement might be interpretively more complex than 'simplify  $expression A$ '. Therefore, procedural complexity and interpretive complexity are in part an outcome of familiarity.

*The zone of low difficulty:* The items in this zone are 2.3.1 and 2.2.2. They are essentially arithmetical in nature and examinees only have to do calculations with numbers. Item 2.3.1 is the easiest. For this item, examinees have to identify that the sequence is arithmetic, find the common difference, identify the first and last term and recall or copy the appropriate formula, do the necessary substitutions and then the calculations. A brute force procedure by writing down all the terms up to 101 and counting the number of terms could also have been followed. These are solution methods that are well-practised over an extensive number of years in different grades. A similar explanation holds for Item 2.2.2. This lends credence to a position that the regular practising over extended periods of particular problems and their solution routines contributes towards success in examinations on these kinds of problems.

*The zone of moderately low difficulty:* The items in this zone are 2.1 and 2.3.2. Item 2.1 deals with the sigma notation as a compact form for representing a sum. Examinees have to do expansion through substitution of a sufficient number of values in order to recognise the kind of sequence, recall or copy the formula for the sum of a geometric sequence and determine and identify the given elements needed for substitution in the formula. It cannot be discounted that the learners could have been made aware of searching for certain cues to identify the sequence for these problem kinds. For example, in this case it might be: 'If the letter indicating the terms is an index of the expression after the summation sign, then the sequence is geometric.' Brute force in terms of 'write down all the terms and add them' can also not be dismissed. As with Item 2.3.1 and Item 2.2.2, once the preliminaries are settled, only arithmetic calculations must be carried out. For these items, there is an increase in procedural complexity.

Item 2.3.2 has an additional interpretive complexity. Meaning has to be assigned to 'even numbers' and even

**TABLE 3:** Clustering of items in zones of difficulty.

Zone of difficulty	Definition	Items
High	More than one standard deviation above the mean	7.2.3
Moderately high	Mean to one standard deviation above the mean	3.1, 3.2, 7.1, 2.2.1, 7.2.1, 7.2.2
Moderately low	Below mean to one standard deviation below the mean	2.1, 2.3.2
Low	More than one standard deviation below the mean	2.2.2, 2.3.1

**TABLE 4:** Item difficulty and Rasch difficulty measures for selected items.

	Item difficulty order											
	Easiest	←									→	Hardest
Item difficulty	2.3.1	2.2.2	2.1	2.3.2†	3.1†	3.2	7.1	7.2.1	2.2.1	7.2.2	7.2.3	
Rasch measure	2.3.1	2.2.2	2.1	2.3.2	3.2	3.1	7.1	2.2.1	7.2.1	7.2.2	7.2.3	
Item number	2.3.1	2.2.2	2.1	2.3.2	3.1	3.2	7.1	7.2.1	2.2.1	7.2.2	7.2.3	
Percentage of sample who did not attempt item	3%	11%	6%	19%	10%	29%	27%	9%	18%	19%	40%	

†These items have the same item difficulty (22%) but the ordering is done according to the Rasch measures.



'remove'. Examinees have to decide which terms will have even numbers and construct a new series with these terms. A challenge is to decide whether it is the terms that are even or the terms in even-numbered positions. It is contended that these complexities of the items place them in a higher difficulty category than those in the zone of low difficulty.

*The zone of moderately high difficulty:* There are six items in this zone. Item 2.2.1 is unusual for this kind of geometric series problem. Normally examinees would not expect to find the value of a variable. Further, the problem calls for deep understanding of a convergent geometric series. To complicate matters (and raise the procedural complexity) it is so that an inequality must be solved. Problem unfamiliarity, procedural and interpretive complexity thus plausibly account for problems of this kind having a moderately high difficulty level.

Item 3.1 requires understanding of a quadratic sequence in terms of second differences. Usual expectation is that second differences will be used to determine a quadratic function. Familiarity with problem type and interpretive complexity account for the item's moderately high level of difficulty. Item 3.2 is contingent on Item 3.1 and the second highest percentage (29%) of the sample cohort did not attempt this item.

For Item 7.1, the number of years is the only concept that is linked to what learners normally encounter. The language used is very compact. The learners must be able to transform the standard formula to accommodate the quarterly interest rate and doubling of an unknown sum. The item is descriptive and general rather than specific in terms of what is given. This item is both procedurally and interpretively complex and 27% of the sample cohort did not attempt it.

That only 9% of the sample cohort did not attempt Item 7.2.1 is indicative that a high level of familiarity with the context inherent in the problem is needed. There are, however, subtleties, delayed payment of the first instalment after the securing of the loan and the number of months that elapsed between 01 February and 01 July, for example, which require a sophisticated level of comprehension to correctly determine the elements needed for problem solution. Furthermore, there is the issue of superfluous information. For the item under discussion the R450 is not required and this goes against the grain of learners expecting that all information must be used. We contend that interpretive complexity accounts for the moderately high level of difficulty of this item.

Interpretive complexity is also at play with Item 7.2.2. It requires that a new principal amount be established as a result of interest accrued up to the end of July. The procedural complexity is also increased because the learners have to make the time period ( $n$ ) the subject in the compound interest formula, which makes it a reversal problem. Learners normally find reversal problems difficult due to limited practice with such problems.

*The zone of high difficulty:* This zone has only one Item, 7.2.3. It is interesting to note that prior questions do not give any scaffolding for dealing with this question and the learner has to start this item afresh. It requires careful interpretation and dealing with sub-problems, which bring procedural complexity into play. These complexities can plausibly account for this item being experienced as the most difficult by examinees, with 40% of the sample cohort not attempting it at all.

## Conclusion

The above findings indicate that the difficulty of items, as evidenced by the performance of the examinees in the high-stakes NSC Mathematics examination, should not only be deduced from the perspective of topics in the curriculum in which learners did not perform satisfactorily. Examinees perform differentially on the sub-questions of a topic. This overall finding echoes that of Mayberry (1983). She found that prospective teachers attained different Van Hiele levels for different quadrilaterals on a test designed to measure the Van Hiele levels of understanding.

From the perspective of classroom teaching and preparing Grade 12 learners for the NSC, the explanatory framework suggests that more attention should be accorded to the development of familiarity through productive practising. Such a focus would contribute tremendously, we contend, to improvement in learners' handling of complex layered questions like question 7.

A final issue that needs careful thought and deliberation is the structure of the Mathematics examination. Common wisdom dictates a structure that is topic based, such as in the case of this article: three questions (2, 3 and 7) dealt with the topic at hand. A further element of the organisation of examination papers is that items perceived as less difficult are placed at the beginning of the subsection. However, the rational choices examiners (including teachers as examiners of in-school designed high-stakes examinations) make might not always parallel those of learners, as exhibited here by the actual responses of learners. For example, Item 2.2.1 was found to be more difficult than Item 2.2.2 by the examinees. This indicates that examiners should give more consideration to item placement in an examination. In the design of a high-stakes examination the assumption is that the placement of less difficult items early in the examination paper will lessen the anxiety load accompanying high-stakes examinations. Consideration should therefore be given to place less difficult items at the start regardless of the topic. Such a reconfiguration might lead to examinees having more courage and willingness to tackle and engage with items higher up a hierarchy of difficulty levels.

The analysis of the difficulty levels of items as manifested by examinees' responses to items in a high-stakes Mathematics examination can provide valuable insights for curriculum interpretation, the design of examinations and for teaching. In the quest for enhanced performance in high-stakes



examinations, thoughtful consideration of the outcomes of such analysis has the potential to positively contribute to this quest. However, such analysis and the dissemination of the findings should be done fairly near to the conclusion of the examination. This will ensure that all major stakeholders can use the outcomes for planning a reasonable time before a next round of teaching for enhanced performance in high-stakes examinations commences.

## Acknowledgements

This work is supported by the National Research Foundation under grant number 77941. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views the National Research Foundation.

## Competing interests

The authors declare that they have no financial or personal interest(s) that may have inappropriately influenced them when writing the article.

## Authors' contributions

C.J. (University of the Western Cape) conceptualised the research and collected the data. With L.H. (University of the Western Cape), he did the quantitative analysis. M.J. (Cape Peninsula University of Technology) and D.M. (University of Cape Town) led the interpretation of the results. The mathematical aspects of the article were attended to by R.L.F. (University of the Western Cape). All authors contributed towards the construction of the final version of the article.

## References

- Backhoff, E., Larrazolo, N., & Rosas, M. (2000). The level of difficulty and discrimination power of the Basic Knowledge and Skills Examination (EXHCOBA). *Revista Electrónica de Investigación Educativa*, 2(1), 1–16. Available from <http://redie.uabc.mx/contenido/vol2no1/contents-backhoff.pdf>
- Carraher, D.W., & Schliemann, A.D. (2007). Early algebra and algebraic reasoning. In F.K. Lester (Ed.), *Second handbook of research in mathematics teaching and learning* (pp. 669–705). Charlotte, NC: Information Age Publishing.

- Cuoco, A. (2005). *Mathematical connections: A companion for teachers and others*. Washington, DC: Mathematical Association of America.
- Department of Basic Education. (2011). *Curriculum and assessment policy statement. Mathematics. Grades 1–3*. Pretoria: DBE. Available from <http://www.education.gov.za/LinkClick.aspx?fileticket=ehGEpQZxz7M%3d&tabid=671&mid=1880>
- Department of Basic Education. (2012). *Curriculum and assessment policy statement. Mathematics. Grades 10–12*. Pretoria: DBE. Available from <http://www.education.gov.za/LinkClick.aspx?fileticket=QPqC7QbX75w%3d&tabid=420&mid=1216>
- Devlin, K. (1996). *Mathematics: The Science of patterns. The search for order in life, mind, and the universe*. New York, NY: Scientific American Library.
- Dunne, T., Long, C., Craig, T., & Venter, E. (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, 16 pages. <http://dx.doi.org/10.4102/pythagoras.v33i3.19>
- Herbert, K., & Brown, B.R. (1997). Patterns as tools for algebraic reasoning. *Teaching Children Mathematics*, 3(6), 340–344.
- Julie, C. (2012). The stability of learners' choices for real-life situations to be used in mathematics. *International Journal of Mathematical Education in Science and Technology*, 44(2), 196–203. <http://dx.doi.org/10.1080/0020739X.2012.703337>
- Julie, C., Leung, A., Thanh, N.C., Posadas, L.S., Sacristan, A.I., & Semenov, A. (2010). Some regional developments in access of digital technologies and ICT. In C. Hoyles, & J.-B. Lagrange (Eds.), *Mathematics education and technology – rethinking the terrain* (pp. 361–83). New York, NY: Springer. [http://dx.doi.org/10.1007/978-1-4419-0146-0\\_17](http://dx.doi.org/10.1007/978-1-4419-0146-0_17)
- Lee, K., Bull, R., Ng, F.S., Pe, L.M., & Ho, R.H.M. (2011). Are patterns important? An investigation of relationships between proficiencies in patterns, computation, executive functioning, and algebraic world problems. *Journal of Educational Psychology*, 103(2), 269–281. <http://dx.doi.org/10.1037/a0023068>
- Linacre, J.M. (2008). WINSTEPS Rasch measurement [Computer software]. Chicago, IL: Winsteps.com.
- Mason, J. 1996. Expressing generality and the roots of algebra. In N. Bednarz, C. Kieran, & L. Lee (Eds.), *Approaches to Algebra: Perspectives for research and teaching* (pp. 65–86). Dordrecht: Kluwer Academic Publishers. [http://dx.doi.org/10.1007/978-94-009-1732-3\\_5](http://dx.doi.org/10.1007/978-94-009-1732-3_5)
- Mayberry, J. (1983). The Van Hiele levels of geometric thought in undergraduate preservice teachers. *Journal for Research in Mathematics Education*, 14(1), 58–69. Available from <http://www.jstor.org/stable/748797>
- Mullis, I.V.S., Martin, M.O., Gonzales, E.J., & Chrostowski, S.J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center. Available from <http://timssandpirls.bc.edu/timss2003i/mathD.html>
- Selter, C. (1996). Doing mathematics while practising skills. In C. van der Boer, & M. Dolk (Eds.), *Modellen, meting en meetkunde. Paradigma's van adaptief onderwijs [Models, measurement and geometry. Paradigms for adaptive education]* (pp. 31–43). Utrecht: Panama/HvU & Freudenthal Institute.
- Thompson, B., & Levitov, J.E. (1985). Using microcomputers to score and evaluate test items. *Collegiate Microcomputer*, 3, 163–168.
- Vogel, R. (2005). Patterns – a fundamental idea of mathematical thinking and learning. *ZDM: The International Journal on Mathematics Education*, 37(5), 445–449. <http://dx.doi.org/10.1007/s11858-005-0035-z>
- Watson, A., & De Geest, E. (2012). Learning coherent mathematics through sequences of microtasks: Making a difference for secondary learners. *International Journal of Science and Mathematics Education*, 10, 213–235. <http://dx.doi.org/10.1007/s10763-011-9290-3>
- Watson, A., & Mason, J. (1998). *Question and prompts for mathematical thinking*. Derby: Association of Teachers of Mathematics.