https://doi.org/10.22364/atee.2022.42

The Development of Numeracy Test Using Three-Dimensional Framework to Assess Numeracy Skills in Grade 7

Ilze France, Marta Mikite, Girts Burgmanis, Dace Namsone

University of Latvia, Latvia
ilze.france@lu.lv; marta.mikite@lu.lv; girts.burgmanis@lu.lv; dace.namsone@lu.lv

ABSTRACT

A growing body of evidence including international level studies (e.g. PISA, TIMSS) demonstrate that numeracy skills (also known internationally by other terms such as mathematical literacy) is crucial for a person's educational achievements and for informed and participatory citizenship. Early and successful interventions to improve students' numeracy skills lie in developing and using valid and reliable diagnostic tests for numeracy skill assessment. This study explored how developing a numeracy test based on three-dimensional framework could be used for numeracy diagnostic purposes in grade 7. To achieve this, initially a three-dimensional numeracy framework based on

1) content knowledge of mathematics,
2) information literacy skills,
3) complexity levels of SOLO taxonomy, was prepared.

Then the framework was used to construct a 32-item numeracy test assessing the ability to use relationships, functions and numerical information in different contexts including science. Next, the instrument was administered to 7th grade students (N = 205) in four schools who were asked to complete 3 hour-long tests. Such diagnostic data could inform teachers on students' numeracy skills and prepare instructional materials that target specific weaknesses in class level. Further, such information could inform personalized student learning instruction and produce improved numeracy diagnostic tests for future use.

Keywords: diagnostic assessment, numeracy skills, test development, test validation, three-dimensional framework, SOLO taxonomy

Introduction

An understanding of mathematics is central to a young person's preparedness for participation in and contribution to modern society. A growing proportion of problems and situations encountered in daily life, including professional contexts, require some level of understanding of mathematics before they can be properly understood and addressed (Organisation for Economic Cooperation and Development [OECD], 2022). One of the key factors in the development of education systems is the availability of increasingly frequent, accurate and detailed feedback mechanisms at different levels of decision making (Csapó & Szendrei, 2011). Tools specifically for diagnosing students' math or science skills, do not measure the students' ability to transfer their math skills in different contexts. It is this transfer that characterizes students' deep thinking and high performance. Such tasks cannot be completed without a degree of cognitive effort, and the underlying conceptual ideas must be engaged with. Higher-level demands mean procedures that require connecting meaning with mathematical operations. A task requires the doing of mathematics when complex thinking is required (Smith & Stein,1998). When mathematical skills are used in different contexts, the term numeracy is used. This study focuses on the development of a reliable and useful diagnostic instrument for measuring students' numeracy skills that is based on a three-dimensional framework. This includes how students cope with various representations of linear functions and solve real life tasks where the mathematical model is a linear function or ratio and proportion in various contexts, including science. The aim of the study is to find out to what extent the different contexts affect students' numeracy performance and what can be concluded about students' ability to use math skills in different situations.

Definition of numeracy

Numeracy is the ability, confidence, and willingness to engage with quantitative or spatial information to make informed decisions in all aspects of daily living (Numeracy Progressions, 2021) Numeracy is not about being able to flexibly use all of mathematics to deal with "life's diverse contexts and situations", but rather to flexibly draw on that subset of mathematics that is most useful in dealing with these "diverse contexts and situations" (Liljedahl & Liu, 2013). Other authors defined numeracy as "ability or tendency to reason critically about quantitative information" (Gittens, 2015). A numerate individual has the confidence and awareness to know when and how to apply quantitative and spatial understandings at home, at school, at work or in the community (Numeracy, 2017). Numerate individuals have "the confidence and competence in using numbers which will allow individuals to solve problems, analyze information and make informed decisions based on calculations'' (Curriculum for

Excellence, 2009). Numeracy is generally seen as some combination of mathematical knowledge, tools, and dispositions, and to be numerate means to be willing and able to use this knowledge, tools, and dispositions across a wide variety of contextual situations (Goos et al., 2013).

A growing body of evidence including international level studies (Programme for International Student Assessment [PISA] or Trends in International Mathematics and Science Study [TIMSS]) demonstrate that numeracy skills are crucial for person's educational achievements and for informed and participatory citizenship. Numeracy skills are also known internationally by other terms such as mathematical literacy. It is more common to use the term numeracy in countries, such as the UK, Canada, South Africa, Australia, and New Zealand. Other names, such as quantitative literacy or mathematical literacy, are used in the USA and elsewhere (Geiger et al., 2015).

In the case of mathematical literacy, PISA is designed to assess if students can make use of their mathematical knowledge in life related contexts as a measure of their readiness for their active participation in society (Geiger et al., 2015). Numeracy emphasizes the use of analysis, inference, interpretation, explanation, evaluation, as well as reflection on one's own reasoning process (metacognition and self-regulation) (Gittens, 2015). The term numeracy assessment is usually used when talking about the assessment of the abilities of adults. The OECD Programme for the International Assessment of Adult Competencies (PIAAC) defines numeracy as the ability to access, use, interpret and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life and numerate behaviour involves managing a situation or solving a problem in a real context, by responding to mathematical content/information/ideas represented in multiple ways (OECD, 2022).

Learner's learning towards conceptual understanding should be ensured through a process of continuous and comprehensive evaluation (Behera, 2021). In the 7th grade most of the number operations have been learned, so it is important to diagnose the ability to use these operations in different contexts before moving on to abstract mathematics. The authors chose the topics "ratio and proportion" and "function" because they are widely used in other contexts, for example in science. Therefore, they meet the criteria to be able to evaluate students' numeracy skills. They fit in one of four key areas of mathematical content, information and ideas that are covered by the numeracy assessment in PIAAC "Pattern, relationships and change" (OECD, 2022). The use of mathematics alone, without a context, cannot be seen as a numeracy activity. Thus, context is at the heart of numeracy (Goos et al., 2019). When the skill of using ratio and proportion and function is used in a scientific context, we can confidently say that numeracy is being demonstrated.

Ratio, proportions and linear function in mathematics and science

One of numeracy's four core dimensions is application of mathematical knowledge (Goos et al., 2019). In turn, teachers are being encouraged more and more to teach the Big Ideas of mathematics: one that links numerous mathematical understandings into a coherent whole. One of the Big Ideas is Proportionality: if two quantities vary proportionally, that relationship can be represented as a linear function (Charles & Carmel, 2005).

In order to be able to diagnose the ability to use ratios and proportions in different contexts, for example in science, it is essential to understand how a student develops understanding in this field. Proportional reasoning, a big idea from mathematics, requires a shift in student thinking away from additive thinking toward multiplicative thinking. This shift is not trivial and unless carefully addressed can lead to student misunderstandings of scale, relation and proportion. (Abramovich & Connell, 2021.) A ratio is a multiplicative comparison of two quantities or measures. A key developmental milestone is the ability of a student to begin to think of a ratio as a distinct entry, different from the two measures that made it up (Van de Walle et al., 2013). Proportion, a big idea from mathematics, is an equation with a ratio on each side. This requires ratio, another big mathematical idea, to be at least partially understood prior to beginning work with proportions (Abramovich & Connell, 2021).

Proportional thinking is developed through activities involving comparing and determining the equivalence of ratios and solving proportions in a wide variety of problem-based contexts and situations without recourse to rules or formulas (Van de Walle et al., 2013). At the simplest level, proportional reasoning, whether involving ratio or proportion, relies on the ability to use multiplicative thinking rather than additive thinking. In other words, instead of describing a relationship between two quantities as being larger by three or smaller by five, the relationship would be described in terms such as triple the size, one fifth the size, four times greater, etc. (Abramovich & Connell, 2021). There are four types of ratios (Abramovich & Connell, 2021, Van de Walle et al., 2013, Musser, et al., 2016.):

1) Part-Part Ratios, where a ratio can relate one part of a whole to another part of the same whole
2) Part-Whole Ratios, which can also be used to represent comparisons of a part to a whole
3) Reflecting Quotients as Ratios can be thought of as a type of quotient
4) Reflecting Rates as Ratios in problems involving students per class, passengers per trolley, etc.

Ratios are extended to understanding and applying proportional reasoning (Van de Walle et al., 2013). Proportional situations are linear situations. Ratios are a special case of linear situations. Proportional quantities are the basis of many main concepts of science, such as speed, pressure, density, composition of mixtures of substances in chemistry, etc. Understanding measurements and unit conversions is crucial to performing any calculation in any field of science.

A three-dimensional framework for numeracy diagnostic instrument

For purposes of the numeracy diagnostic assessment, a three-dimensional framework has been developed (Burgmanis et al., 2021), which has foundations in several prominent previous studies and frameworks. The authors suggest that numeracy can be analyzed in terms of three inter-related dimensions:

1) Content of mathematical knowledge important for designing of the items;
2) Information literacy skills important to apply mathematical knowledge and skills in various contexts;
3) The complexity of student's performance to analyze and provide evidence on the structure of student's understandings and application of skills.

The first dimension of the framework "mathematical content" includes following categories:

1) quantity and numbers
2) ratio and relationship
3) data analysis and probability
4) measurement
5) space and shape
6) location and directions.

The second dimension of the framework is information literacy skills, which refer to ability of student to acquire, analyze, utilize and communicate mathematical knowledge in various contexts. The third dimension of the framework is complexity of student's performance. The authors of the framework use Structure of the Observed Learning Outcome (SOLO) taxonomy (Biggs & Collis, 1989).

Methodology

Participants

The diagnostic assessment instrument was administered using a sample of 205 7th grade pupils from four schools in Latvia. The sample of the study was selected using convenience sampling which is widely employed in the pilot test of an under-developed instrument. All pupils in 7th grade who attended the school on the days when the pilot tests were given completed the test and were included in the study sample. The sample of schools were representative and stratified by mathematics achievements (i.e., high, above medium, below medium and low achievement) of pupils from 9th grade in national level testing. Agreements were signed between the Rector of the University of Latvia, the municipality and the school, setting out the procedures for collecting and processing data for the study. This agreement was not violated during the study and the ethical standards of the study were respected.

Procedure

The study was carried out in two stages:

1) developing the numeracy skills tests based on a three-dimensional assessment framework that includes constructing items that target the specified attributes,
2) statistical analysis of data from test administration using item analysis and Rasch analysis.

The first stage of the study was to carry out an extensive study on the literature pertaining to diagnostic assessment, measuring numeracy skills and test development for the purpose of developing a numeracy skills tests based on a three-dimensional framework. The second phase included test administration in schools.

Instrument and item development

A three-dimensional framework was used to construct a 32-item numeracy test focused on the context category "ratio and relationship". It was divided into three parts:

1) Functions,
2) Ratio and Proportion,
3) Ratio and relationship in science.

The test items were designed to correspond to specific content of mathematical knowledge sections (Table 1). Correspondence of items in information literacy and complexity of students' performance is summarized in Table 2.

Table 1. Structure of the test: item correspondence with content of mathematical knowledge

The skill used in the item	Item
Use of different types of ratios	S_2.1.; S_2.2.; S_3.1.; S_3.2.; R_1.1.; R_1.2.; R_1.3.; R_1.4.; R_1.5.; R_1.6.; R_1. 7.; R_2.1.; R_2.2.;
Determine the unknown term of the proportion	S_1.2.; F_1.; R_3.1.; R_3.2.; R_3.3.;
Sees the relationship using a graph	S_1.1.; S_3.3.; F_4.1.; F_4.2.; F_4.3.; F_4.4.; F_5.;
Represent the relationship graphically	F_6.1.; F_6.2.; F_7_1.; F_7_2.;
See the relationship using a table or formula	F_2.; F_3.; F_7.3.;

Note. The name of each item is a code in which the letters represent the test in which it is included (F – Functions, R – Ratio and Proportion, S – Ratio and relationship in science), and the numbers represent the item number in the test.

Table 2. Structure of the test: item correspondence to information literacy skills and level of performance complexity

	Unistructural	Multistructural	Relational and Extended Abstract
Acquire information	S_1.1.	S_3.3.; R_3.1.; F_4.1.; F_5.	F_4.2.; R_1.7.
Analyze information	R_1.1.; R_1.4.	S_3.1.; R_1.2.; F_1.; F_6.1.; F_2.; S_2.1.	R_1.6.; F_3.; F_7.3.; S_1.2.; R_3.2.; F_6.2.; F_4.3.; R_1.5; R_2.2.;
Utilize information and communicate	F_7.2.	R_1.3.	R_2.1.; S_3.2.; S_2.2.; R_3.3.; F_7.1.; F_4.4.;

The first two tests in mathematics and the third test in science were created based on the same framework with two groups of experts working separately. Each group of experts designed items, approbated in smaller groups of students and improved the tests during the development process. Next, the instrument was administered to students in grade 7 ($N = 205$) in four schools who were asked to complete the tests in 3 series each 1-hr long.

Data analysis

In this study Item Response Theory (IRT) was used to accurately examine a set of items that is measured. IRT is a collection of measurement models that attempt to explain the connection between observed item responses on a scale and an underlying construct (Cappelleri et al., 2014). Content validity was established by 3 experts. The Delphi methodology was used. Experts did a quantitative (scale 1–5) evaluation of items using two criteria: (1) item coherence with indicator, (2) item correspondence to specific level of SOLO taxonomy. The student completed tests were also evaluated by experts. Assessment calibration was done using the following procedure. 20 works were initially selected from each test and evaluated by several experts. The results were compared, and the reliability of the correction was evaluated. To evaluate the quality of items a Rasch model for the Partial Credit Model (Masters, 1982) was used. Previous studies show that Rasch techniques can be used to document and evaluate the measurement functioning of diagnostic instruments through analysis of item fit (Boone et al., 2013). For the statistical analysis the SPSS software version 26.0 and WINSTEPS version 4.6.2 (Linacre, 2015) was used.

Results

Students' performance in developed numeracy tests

The student results in each item were analyzed using a Rasch Wright Map (Figure 1). Since the item mean is one standard deviation higher than the person mean, it can be assured that the items were difficult and challenging for this group of students. Items above the mean value are more difficult than the ones below.

```
MEASURE         Person - MAP - Item
                         <more>|<rare>
      2                        +  F_3       R_1_6     S_3_3
                               |
                               |
                               |
                               |
                               |S F_7_3     R_2_1
                               |
      1                       T+  R_3_2     S_1_2     S_3_2
                           .#  |  R_1_3     S_3_1
                            .  |  F_4_2
                          .##  |  F_6_2     R_1_7     S_2_2
                        #####  |  F_4_3     R_1_2     R_3_1
                          ###  |  R_3_3
                        ##### S|M F_1
      0                  .###  +  F_7_1
                    .########  |  R_1_5
                   .#########  |  F_4_4     R_1_1
                        #####  |  F_6_1     R_1_4     R_2_2
                       ######  |
                    .######## M|
                         ####  |S
     -1                 #####  +
                    .########  |  F_7_2     S_1_1
                       ######  |
                       .#####  |  F_2
                           .#  |  F_4_1
                              S|  F_5
                           ##  |T
     -2                    .#  +  S_2_1
                               |
                           ##  |
                               |
                       .##### T|
                               |
     -3                        +
                          .##  |
                               |
                               |
                               |
                               |
                            .  |
     -4                        +
                         <less>|<freq>
EACH "#" IS 2: EACH "." IS 1
```

Figure 1. Rasch Wright map

Students’ numeracy performance in different contexts

The authors use the following procedure to select the items to be studied further. Three items in the science test have been selected. To cover a larger part of the construct, tasks are from different categories of information literacy (Table 2). Then two corresponding items from mathematics were selected for each of three science items, creating three sets of three items from the same construct field and the same mathematical skill used to complete the item. From the obtained data, students’ ability to use mathematical skills in different contexts can be ascertained. In order to see how students make a transfer between different contexts, students’ performance in selected items was compared using SOLO taxonomy where it was possible.

Acquiring information in various contexts

Analyzing the students' ability to obtain information, 3 tasks with the same level of performance complexity were selected for deeper research. In item S_3.3. students select the appropriate relationship graph after being given a precise description in a scientific context. Students need to determine the direction of a graph of a linear function to match the given situation. A similar item is F_5. Only here students need to choose a schedule based on the values given in the description of the situation. In the third task F_4.4. the graph shows the path of a girl going home, where the distance depends on time. Students must be able to read from the graph where the halfway point is and in how many minutes the girl will reach it. In the first item, the context is considered scientific, in the other two, mathematical. Student performance in a science context differed significantly from using the same skill in a math context (Table 3).

Table 3. Students results of acquiring information in different contexts

Item and context	S_3.3. (Science context)	F_4.1. (Mathematical context)	F_5 (Mathematical context)
Correct answers, percentage	8%	65%	68%

Analyze information in various contexts

In the first item (S_1.2.) of this construct field, students have to use percentages in the calculation of proportional quantities. In the second item (R_3.2.), they have to see the relationship that connects the amount of fuel consumed and the number of kilometres travelled. Information must be obtained from the graph in the third item (F_4.3.) about the relationship between the distance travelled and the time spent, which is the velocity. The first item's context is scientific, the second is real- life and the third is mathematical. The maximum level of performance complexity in all items is multistructural. Collected data on the percentage of students whose outcome is at the unistructural and multistructural level is shown in Table 4. The performance of the other students can be evaluated at the pre-structural level.

Table 4. Students results of acquiring information in different contexts

Item and context	Unistructural	Multistructural
S_1.2 (Science context)	10%	5%
R_3.2. (Real-life context)	23%	2%
F_4.3. (Mathematical context)	12%	18%

The results show that in the contexts of science and everyday life, most of the students operate at the unistructural level, which is different from the context of mathematics, where the number of students with answers at the multistructural level is significantly higher. This indicates that students have problems transferring between different contexts.

Communication in different contexts

In the following three items, the task is to mathematically justify a statement. In the science item, students must "explain choice using calculations". In item R_3.3. which corresponds to the real-life context, students need to write a "recommendation based on calculations of how much fuel Monika needs to fill up when going to visit her sister". In F_4.4. students have to describe a graph in order to justify the falsity of the statement.

It can be seen from Table 5 that students show low performance on relational and extended abstract level in all contexts. In the context of science most of the students performed at the unistructural level. In the context of mathematics, students performed relatively well at the multistructural level, but did not make the transfer to the relational and extended abstract level.

Table 5. Students results of communication in different contexts

Item and context	Unistructural	Multistructural	Relational and Extended Abstract
S_2.2. (Science context)	34%	8%	4%
R_3.3. (Real-life context)	19%	17%	8%
F_4.4. (Mathematical context)	19%	40%	5%

Discussion

Low students' performance in numeracy tests may have different reasons. To understand what those might be, it is necessary to analyze the experience of testing numeracy in Latvia so far, as well as the results of research in other countries. There is also a need to consider how numeracy is taught in schools.

The lack of pragmatic clarity as to what numeracy is, coupled with a lack of resources around this important construct, afforded the emergence of a more intuitive and grounded entry into numeracy (Liljedahl, 2015). For students to become numerate, they must engage with tasks that demand the use of mathematics in multiple contexts, and so effective numeracy instruction must take place in all school subjects, not just mathematics (Steen, 2001). It is therefore particularly important to measure the same construct in multiple subjects, as the three-dimensional framework is measured in mathematics and science in this study.

Reason for the low results could be that no separate measurements of students' numeracy skills have been carried out in Latvia so far. When initially performing such measurements, students' results tend to be low (Liljedahl, 2015). In a previous study (Cao et al., 2022) on students' performance in numeracy in Vietnam one of the findings was low results on ability to apply mathematics knowledge and skills in everyday situations. Referring to authors (Cao et al., 2022) a significant reason for these results is the fact that students were still learning mathematics according to the old curriculum that was mostly based on mathematics content rather than the competency-based approach. Before implementing the new curriculum in Latvia, more attention in lessons and tests was paid to obtaining answers, not on how answers were obtained, decision and conclusions based on the obtained results.

This type of assessment allows feedback on performance at a particular point in time but does not give the opportunity to monitor students' progress. Norwegian researchers (Ræder et al., 2022) implement a vertical scaling design for existing assessment system. To keep track of how numeracy skills are developing, they use ongoing national assessments. Inspired by this further research could examine the possibility of linking the three-dimensional framework and national assessments in mathematics and science, looking for ways to track students' progress.

Conclusions

Analyzing the performance of the students in all three tests in general, using the Rasch Wright map, it can be concluded that this diagnostic tool was very challenging for the students and the performance of the students is low. No unequivocal relationship was observed between students' performance on the tasks and the cognitive depth of the tasks.

The three created tests, combining tasks in mathematics and science, give an opportunity to determine students' numeracy skills. They provide an opportunity to verify to what extent students can perform mathematical operations, and to what extent they can build the transfer of knowledge and skills in different contexts. The performance of the students included in the study in the contexts of science and real-life is lower than when using the same mathematical skills in the context of mathematics.

In any context, students demonstrate poor performance at the Relational and Extended Abstract level, which indicates the need to analyze the learning process and place greater emphasis on tasks with higher performance complexity.

Aknowledgment

This research was supported within the project "IT-based support system prototype for providing feedback and improve student performance in literacy and numeracy acquisition" Agreement no. 1.1.1.1/19/A/076

REFERENCES

Abramovich, S., & Connell, M. L. (2021). *Developing Deep Knowledge in Middle School Mathematics*. Springer International Publishing.

Behera, B. (2021). Does Diagnostic Evaluation in Mathematics Improve Learners' Learning? *Journal of Global Research in Education and Social Science, 15*(2), 1–8.

Biggs, J., & Collis, K. (1989). Towards a model of school-based curriculum development and assessment using the SOLO taxonomy. *Australian journal of education, 33*(2), 151–163.

Boone, W. J., Staver, J. R., & Yale, M. S. (2013). *Rasch analysis in the human sciences*. Springer Science & Business Media.

Burgmanis, G., France, I., Namsone, D., & Čakāne, L. (2021). Development and Validation of Diagnostic Assessment Instrument for Numeracy Skills in 7th Grade. In *ICERI2021 Proceedings* (pp. 7781–7791). IATED.

Cao, T. H., Nguyen, H. C., Dang, X. C., Chu, C. T., Le, T. A., & Le, T. T. H. (2022). Exploring Numeracy Skills of Lower Secondary School Students in Mountainous Areas of Northern Vietnam. *International Journal of Learning, Teaching and Educational Research, 21*(10).

Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical therapeutics, 36*(5), 648–662.

Charles, R. I., & Carmel, C. A. (2005). Big ideas and understandings as the foundation for elementary and middle school mathematics. *Journal of Mathematics Education, 7*(3), 9–24.

Csapó, B., & Szendrei, M. (2011). *Framework for diagnostic assessment of mathematics*. Nemzeti Tankönyvkiadó.

Curriculum for Excellence, Scottish Government (2009). https://education.gov.scot/documents/All-experiencesoutcomes18.pdf

Dole, S., & Geiger, V. (2020). *Numeracy across the curriculum: Research-based strategies for enhancing teaching and learning*. Routledge.

Geiger, V., Goos, M., & Forgasz, H. (2015). A rich interpretation of numeracy for the 21st century: A survey of the state of the field. *ZDM, 47*, 531–548.

Gittens, C. A. (2015). Assessing numeracy in the upper elementary and middle school years. *Numeracy, 8*(1), 15–28.

Goos, M., Geiger, V., & Dole, S. (2013). Designing rich numeracy tasks. *ICMI Study 22: Task Design in Mathematics*, 589–597.

Goos, M., Geiger, V., Forgasz, H., Bennison, A., & Dole, S. (2019). *Numeracy Across the Curriculum* (1st ed.). Taylor and Francis.

Liljedahl, P. (2015). Numeracy task design: A case of changing mathematics teaching practice. *ZDM, 47*, 625–637.

Liljedahl, P., & Liu, M. (2013). Numeracy. *Vector, 2*, 34–39. http://www.peterliljedahl.com/wp-content/uploads/NR-Numeracy.pdf

Linacre, J. M. (2015). *Winsteps® Rasch measurement computer program*. Winsteps.com, Beaverton.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174.

Musser, G. L., Burger, W. F. & Peterson, B. E. (2016). *Mathematics for elementary teachers: A contemporary approach*.

Numeracy, Goverment of Alberta (2017). https://education.alberta.ca/literacy-and-numeracy/numeracy/everyone/numeracy-resource-links/

Numeracy Progressions, Govermet of Alberta (2021). https://curriculum.learnalberta.ca/cdn/progression-pdfs/Numeracy%20Progressions%20-%20EN.pdf

OECD, PISA, O. (2022). Mathematics Framework (Draft). *PISA*. https://pisa2022-maths.oecd.org/files/PISA%202022%20Mathematics%20Framework%20Draft.pdf

Ræder, H. G., Andersson, B., & Olsen, R. V. (2022). Numeracy across grades–vertically scaling the Norwegian national numeracy tests. *Assessment in Education: Principles, Policy & Practice*, 1–21.

Smith, M. S., & Stein, M. K. (1998). Reflections on practice: Selecting and creating mathematical tasks: From research to practice. *Mathematics teaching in the middle school*, *3*(5), 344–350.

Steen, L. A. (Ed.). (2001). *Mathematics and democracy: The case for quantitative literacy*. Princeton, NJ: NCED.

Van de Walle, J. A., Karp, K. S., & Bay-Williams, J. M. (2016). *Elementary and middle school mathematics*. London: Pearson Education UK.