EduMa: Mathematics Education Learning And Teaching
December 2023, Vol 12 No 2 Page 259 – 272
https://syekhnurjati.ac.id/jurnal/index.php/eduma/eduma/article/view/13318
p-ISSN: 2086-3918, e-ISSN: 2502-5209
Published by Tadris Matematika IAIN Syekh Nurjati cirebon

EduMa
MATHEMATICS EDUCATION LEARNING AND TEACHING

Analysis of Students' Ability and Multiple-Choice Numeracy Test based on Item Difficulty Using Rasch

Churun Lu'lu'il Maknun[1*], Ajeng Wahyuni[2], Irwani Zawawi [3]

1 Universitas Muhammadiyah Gresik, Mathematics Education
2 Institut Agama Islam Negeri Ponorogo, Faculty of Econimic and Business
3 Universitas Muhammadiyah Gresik, Mathematics Education

*Corresponding author: Jalan Sumatra No 101 , Gresik, Jawa Timur, 61121, Indonesia. e-mail addresses: churunlm@umg.ac.id

article info

How to cite this article:

Maknun, C.L., Wahyuni, A., & Zawawi, I. (2023). Analysis of Students' Ability and Multiple-Choice Numeracy Test Based on Item Difficulty Using Rasch. Eduma : Mathematics Education Learning And Teaching, 12(2), 259 - 272. doi: http://dx.doi.org/10.24235/eduma.v12i2.13318

Article history:

Received: 03 30, 23

Accepted: 10 20, 23

Published: 12, 23

abstract

The purpose of this research is to evaluate learners' numeracy knowledge. This study looked into how university students were taught math skills. The research strategy used in this investigation was quantitative. Rasch Model study was used for the data study. The first research question's goal was to assess the numeracy exams' item difficulty characteristics using the Rasch Measurement Model, and The second research topic's goal was to look into the student's numeracy test performance parameters. According to the study, performing a range of calculations to add, subtract, and multiply as well as calculating and interpreting mean in a variety of contexts were the skills that students found the most difficult to master. The pupils perform exceptionally well when they identify and contrast information and data in tables. The outcome makes it possible to pinpoint each examinee's testing assets and weaknesses; this diagnostic should be applied to raise students' numeracy standards. The level of numeracy can be improved using two important methods. Teaching strategy: Math instructors should adapt their course materials to the students' learning styles

Keywords:
Numeracy, Item Difficulty, Rasch Analysis, Test

INTRODUCTION

Numerous academic fields at the institution require mathematics expertise. Particularly, degrees such as Engineering, Nursing (Gregory et al., 2019), and Economics have highlighted the mathematical abilities required for students to become specialists in their respective fields. Concern exists that these abilities are inadequate for the quantitative culture of today. Staff in academic assistance or service programs must have a thorough understanding of this metaknowledge concerning the various mathematics skills (Galligan, 2013). Nevertheless, it is equally crucial to establish an institutional structure around the production of such knowledge. The article will begin by defining numeracy. The report will provide a result of numarcy ckill portarit at the university level.

Numeracy is the capacity to access, utilize, comprehend, and convey mathematical information and concepts such as numbers, facts, and mathematical symbols. (OECD, 2016). This ability may be utilised to participate in and manage mathematical demands and to assist in solving the challenges of human and adult life in the job and in daily life (Jonas, 2018). Reading bills, invoices, bank statements, or other financial statements, using a calculator, and calculating prices, expenses, or budgets are the most often stated numeracy activities. Unhappily, a considerable majority of adults in Jakarta (Indonesia) have inadequate numeracy abilities, which are limited to counting, sorting, completing basic arithmetic operations with whole numbers or money, and recognising common spatial representations (OECD, 2015). The relationship between numeracy and salary indicates that more competence is rewarded with higher compensation.(Vignoles, 2016; Windisch, 2015). Adults aged 16-34 with inadequate numeracy abilities were three times more likely to be unemployed than those with adequate skills. Even when working, persons with inadequate numeracy abilities were frequently locked in precarious, low-paying positions and were less likely to get work-related training or promotion (Windisch, 2015, 2016). Adults with low levels of education may need to acquire new skills to succeed in an evolving workplace or a new vocation. Without good literacy and numeracy abilities, however, learning new skills is difficult. The programme that improves an adult's proficiency is difficult to access, particularly for individuals older than 45. (Windisch, 2015). Thus, the sooner numeracy skills are identified, the simpler it is to get the numeracy skill competency, which may be determined at school and university..

The low numeracy literacy skills in Indonesia, especially for students, are known from the PISA (2018) results in mathematics and TIMSS (2016) tests. Despite the fact that numeracy and math competency are not identical, both are founded on the same knowledge and abilities. The distinction is in the application of this information and abilities. Numeracy is the ability to obtain, utilize, comprehend, and convey mathematical knowledge and ideas, concepts, and principles in real-world contexts. PISA gave Indonesia a math score of 379 out of an average score of 489, whilst TIMSS gave Indonesia a math score of 395 out of an average score of 500. Based on these outcomes, Indonesia holds the bottom spot. However, there is a shortage of information about undergraduate students' numeracy ability. In contrast, higher education must identify the students who are most susceptible to failing to meet the curriculum's numeracy requirements in order to make the most equitable use of intervention resources. Therefore, institutions of higher education must design mechanisms that properly identify students most deserving of further help (Prince & Frith, 2020).

This study helps to universities' student-related decision-making in the following ways:

- Students must be able to combine their abilities from school and the workforce (if appropriate) into their university courses in order to critically analyse and assess their own and others' work in order to be numerate.

- To strengthen this skill, pupils must be able to reflect on school mathematics and adapt it to the objective, anticipate workplace mathematics, and analyse critically the contextual relationships between variables.

Being numerate is a one-of-a-kind chance for students to view mathematics in this better perspective – to incubate, think on, and improve mathematics learnt in school, which frequently goes frozen or underused in the workplace (Galligan, 2013).

This research investigated the numeracy skill instruction of student unversity. The primary emphasis is on numeracy in regard to its educational relationship with literacy. Concern for the development of general skills is reflected in the educational concepts of the university.

LITERATURE REVIEW/ THEORETICAL FRAMEWORKS

Numeracy Skill

In practice, the phrase may refer to a variety of concepts, such as computational arithmetic, necessary mathematics, social mathematics, daily survival skills, quantitative literacy, mathematical literacy, and a facet of mathematical ability. These descriptions cover a range of human qualities, from fundamental skills to advanced cognitive abilities such as problem solving and communication. Academic numeracy includes: mathematical expertise in the specific context of a career field and the academic reflection of the profession at the time, greater understanding of the mathematics in the context and in students' own mathematical knowledge, requiring both cognitive and metacognitive skills, and confidence – highlighting its profoundly affective nature. These three elements are crucial for directing a plan to assist kids become more numerate. (Galligan, 2013).

The Australian curriculum defines general capabilites for numeracy i..e

- Using measurement
- Estimating and calculating using the whole number
- Recognising and using pattern and relationship
- Using fraction, decimals, percentage, ratios and rates
- Using spatial reasoning
- Interpreting statistical information

The NSW Department of Education Standart of Numeracy Skill

- Mental computation and numerical reasoning
- Pattern and algebra
- Spatial visualitation, geometric reasoning and mapping
- Measurements and time calculations
- Statistics and probability

Mental computation and numerical reasoning

Without the use of paper, a calculator, or any other device, the process of solving a problem in one's head and arriving at accurate or approximative results mentally is referred to as mental computing (Jordan, Glutting, & Ramineni, 2010). Children need to learn how to mentally compute, but the emphasis shouldn't just be on teaching them computation strategies; it should also be on teaching them higher order thinking, reasoning, and critiquing skills, as well as how to understand numbers and number operations (Erdem & Gürbüz, 2014). It is a crucial way for kids to think since it helps them develop deeper understanding of how numbers connect to one another, decision-making skills, and calculation strategies (Tsao, 2011; Everett, Harsy, Hupp, & Jewell, 2014). In daily life, mental calculation is the most prevalent type of computing. For rapid calculations and estimations, it is utilized. Classroom instruction emphasizes mental computation, particularly the development of students' abilities to compute additions, subtractions, multiplications, and divisions mentally (Singh et al., 2020). The diversity in financial literacy was significantly explained by mental arithmetic, math proficiency, and numerical reasoning (Sunderaraman, Barker, Chapman, & Cosentino, 2022)

Pattern and Algebra

Exploration of patterns is a fundamental aspect of all mathematical and scientific disciplines. Children who are attempting to represent apparent patterns numerically are in an ideal position to learn algebraic language and engage in algebraic activities. By includes pattern work, the Algebra component of the Principles and Standards for School Mathematics recognises the relationship between pattern discovery and algebraic reasoning (Lee & Freiman, 2006). While algebraic thinking is a wide phrase that incorporates the variety of reasoning and modalities of representation utilised in algebraic activities, the term algebraic reasoning is more specific. While there are some differences, the majority of frameworks identify comparable activities and abilities, for instance, early algebra is described by Kaput (2008) in terms of three content strands: i) the study of structures, ii) the study of functions, and iii) the application of a cluster of modelling languages within and outside of mathematics. (Bråting & Kilhamn, 2021).

Spatial visualitation, geometric reasoning and mapping

Shapes, locations, pathways, connections between entities, and relationships between entities and reference frames comprise spatial information. This knowledge is represented in human cognition and may be cognitively changed to aid physical world manipulation, building, and navigation, as well as scholastic and intellectual achievement (Newcombe & Shipley, 2015). According to Hegarty & Waller (2004), spatial orienting ability is the capability to see oneself or an arrangement from numerous viewpoints. Working with maps requires understanding how to orient them in space; this requires perception. There are three components to spatial reasoning: (a) ideas of space (e.g. distance between objects, distribution and specific position of things), (b) instruments of representation (e.g. maps, graphs, GIS), and (c) procedures of reasoning (different ways to reason about a geographical scenario) (Scholz, Huynh, Brysch, & Scholz, 2014). Map readers use spatial thinking to interpret the relief in a spatial scenario from a map (a) with various tools of representation (b), creating a mental representation (c) of topographic information recognizing the two or three-dimensional shapes, structures, positions, and orientations of objects (Jarvis, 2011). Another evidence shows that males enhanced their spatial orientation ability more than females by genderMap readers analyse the relief in a spatial situation from a map (a) using various instruments of representation (b), constructing a mental representation (c) of topographic information by recognising the two- or three-dimensional forms, structures, locations, and orientations of objects. (Jarvis, 2011). Males improved their spatial orientation skill more than females, according to further findings. (Carbonell Carrera & Bermejo Asensio, 2017).

Measurements and time calculations

Numeracy includes fundamental computation, but also the ability to comprehend measurement, estimation, and logic, to perform multistep operations, and to discern which mathematical concepts must be applied to interpret a situation and then solve problems (Rothman et al., 2006). The objective numerical questions assess a candidate's ability to calculate risk magnitudes using percentages and proportions, and to convert percentages to proportions, proportions to percentages, and probabilities to proportions. (Fagerlin et al., 2007).

Statistics and probability

Situations requiring numeracy may involve numbers, quantitative or measurable data, or visual or textual data based on mathematical principles or including mathematical aspects. All of the following circumstances are relevant to a discussion of probability literacy: calculations, interpretations, and judgements (Gal, 2005). robability is implicitly and overtly incorporated into an extensive variety of real-world events and processes. Adults must be competent in circumstances demanding the understanding or creation of probabilistic messages or decision-making. Further Gal (2005) provided a model suggesting that effective interaction with probability-laden circumstances necessitates the existence of five knowledge-bases and supporting attitudes. (concerning big concepts, calculating probabilities, language, context, and critical questions). These parts of probability literacy are embedded within a wider landscape of vital competences or fundamental abilities, which are already described by broad and interconnected notions such as reading, numeracy, and statistical literacy, as well as scientific literacy and health literacy.

METHODS

Participants were 379 subjects collected from two separate samples of university students. The first sample consisted of 288 junior year college members of the economics program. The second sample consisted of 91 sophomore year college members of the banking program. Both of the participants were students at the faculty of economics and business who voluntarily participated in the survey. These two sample groups were merged into a single dataset. The sample included 304 women (80.2%) and 75 men (19.8%). The range of age for this sample was 19-21 years old.

Data Collection and Analysis

The data in this study were collected through an online test distributed to the participants. The participants could only fill the form once and could not be edited after submitting the responses. There was no limitation and particular time in response to the test so that the participants could fill the test at their earliest convenience. Once the data had been collected, instruments analyses such as validity, reliability, and item difficulty were conducted to investigate which items were accepted for further investigations. Thus, the numerical analysis in this study relied on the things which pass the validity and reliability, not the whole items given to the participants. The items were analyzed using SPSS for validity, reliability, descriptive statistics, and comparative study. The objective numeracy score was determined by tallying the number of right responses. The left-blank responses were considered wrong. (as long as the respondent answered any questions at all).

Instrumentation

The instrument in this study aims to investigate numeracy among university students. It was derived with some modifications from the minimum standard of numeracy sample test of the NSW Departement of Education (NSW-Government, n.d.). It contains 24 multiple-choice items, which cover five focus areas. There were four possible scores of responses that only had one correct answer in all the items measuring the constructs in the numeracy instrument. The purpose for adapting items from sample tests of the NSW Department of Education is because of their suitability in the context of the participant of this study. The numeracy skill test provided by the NSW Department of Education was intended for students who intend to leave high school, approximately the same age as the participant in this study. However, in some cases, the items provided had international context and measurement, which is less contextual for this study. Despite this, some of the items were rematched with Indonesian characteristics.

RESULT AND DISCUSSION

Validity Test

This study adheres to the authoritative recommendations of AERA et al. (1999) regarding sources of validity evidence, which are established through the presentation of accumulated evidence based on test content, response processes, internal structure, relationship to other variables, and consequences of testing. The findings from internal consistency reliability and factor analysis are regarded as evidence of internal structure. Factor analysis can give statistical evidence of the degree to which responses adhere to the stated items. In addition, exploratory factor analysis is used in this study to determine the number of test factors. Also, it is one of the most useful strategies in instrument development for proving internal structure-based validity (Henson, Capraro, & Capraro, 2004; Kieffer, 1999).

Factor analysis is a helpful technique for establishing valid evidence based on an internal structure in instrument development (Wetzel, 2011). Since the outcome of the data was a dichotomy, thus KR-20 Cronbach Alpha was used to prove the internal consistency reliability. The factor analysis and KR-20 Cronbach Alpha used SPSS to compute the data. The result of the factor analysis is provided in Table 1. The KMO measure of sampling adequacy was 0.84; sig. 0.0, which is more than 0.5, means the number of samples is adequate to proceed with the factor analysis test (N.A., 2019). The sample used in factor analysis is 379 as Tabachnick and Fidell (2007) recommendation of at least 300 participants. Since this study deals with population for other solutions with few indicators per factor and over-extraction, the method of extraction uses Principal axis factoring (PAF) and maximum likelihood factor analysis (MLFA) (De-Winter & Dodou, 2012) . Varimax rotation method, and 0.3 of factor loading. The first run of analysis factors led to comprising of 7 factors. However, the 3rd, 4th, 5th, 6th, and 7th did not have three items loading more than 0.4 in the rotated factor matrix, so they were excluded, and the analysis was re-run to extract two factors only (Samuels, 2016). The items loading on each factor were noted as Table 1.

Table 1
The distribution of items based on the genarating factors

Factor	Items
1	5, 8, 11, 12, 13, 15, 16, 17, 18, 19, 20, 22, 23, 24
2	1, 2, 4, 6, 7, 9, 10, 14

Further analysis was conducted to ensure that the items were in the proper factor classification. Test criterion relationships could show the correlation between factors and between factors (Wetzel, 2011). In this study, the Pearson Product Moment was used with the standard error of 0.05. The average correlation in factor 1 was 2.32, then the moderate correlation of factor 2 was 1.76. Thus the overall correlation within the factor was 2.04. The Pearson correlation between factors was 0.33. This value was acceptable as the within-group correlation should have been considered higher. Finally, the reliability check for the factors was run. Reliability is concerned with the consistency, stability, and dependability of an assessment or questionnaire (McMillan, 2007). Under the conceptualization of validity as a unitary concept, reliability is understood to support validity based on internal structure and response process. The Cronbach's alpha of factors 1 and 2 are 0.75 and 0.68, respectively. These reliability scores for both factors were fair to conduct further analysis(Mohamad, Lisa, Sern, & Mohd, 2015). The item difficulty was retrieved by calculating the ratio of who answered an item correctly and total item tested (McCowan & McCowan, 1999). Table 2 shows the distribution of items in the instrument.

Table 2
Factor 1 of Item distribution

Difficulty	Number	Question
0.84	5	A kid departs for school at 7:40 am and arrives at 8:15 am. How long does his commute to school take?
0.87	8	The image depicts a ruler measuring a pencil. What is the pencil's length?
0.50	11	A painter combines blue paint with white paint in a ratio of 4:1 to create a hue of blue. How many litres of blue and white paint are necessary to create 60 litres of the desired shade of blue?
0.69	12	It is believed that a person's height (H) in centimetres equals 4 times the length of their forearm (F) plus 10 centimetres. This might be expressed as a formula H = 4F + 10. According to the calculation, what is Ben's predicted forearm length if he is 170 cm tall?
0.77	13	Each minute, 100 litres of water are drawn from a tank to irrigate fields. After 10 minutes of watering the plants, the tank filled 8,000 L. How much water was in the tank prior to watering the crops?
0.49	15	The graphic depicts a portion of a home roof layout. What is the angle's magnitude at X?
0.77	16	A airliner had 280 people on board. The majority of passengers were adults. How many adults were on board the aircraft?
0.62	17	The Jumping Jeans store is holding a sale. The table displays regular and discount pricing. Which denim brand's standard pricing has been decreased by one-third?
0.51	18	Glenn gets paid $40 per hour, but on Saturdays he receives double pay. Monday through Friday, he works six hours a day, and on Saturday, he works four. How much does he earn every week?
0.73	19	The mass of a horse is 0.6 tonnes, whereas the rider weighs 53 kilogrammes. What is the combined weight of the horse and the rider?
0.40	20	Shastri is employed by a neighbourhood store. A client pays for her purchase with a $50 bill. Three dollars per kilogramme for 500 grammes of tomatoes, seven dollars per kilogramme for 100 grammes of mushrooms, and ten dollars per kilogramme for 250 grammes of strawberries. What is the proper amount of change for Shastri to return to the customer?
0.40	22	Michelle's average score on three tests is 70. What needs she achieve in her upcoming exam to raise her average to 74?
0.48	23	Alex is going between towns A and B. Her map has a 1:100,000 scale. She measures the distance between the two cities using a ruler. On the map, the distance is 2,6 cm. How far must Alex travel?
0.51	24	A wheelbarrow weighs w kg when empty. It is filled with n bricks each weighing b kg. The total weight in kilograms of the wheelbarrow and bricks is

Table 3
Factor 2 of Item Distribution

Difficulty	No	Question
0.91	1	Who has the highest score?
0.86	2	Which sign has the shape of an octagon?
0.91	4	What is 20% of $25.00?

0.88	6	Which hotel should he book?
0.81	7	Kim just got off the bus at the Bus Station on Main St. In which direction does Kim have to walk to get to the County Hospital?
0.92	9	6 five-dollar notes • 1 ten-dollar note • 2 fifty-dollar notes. How much do I have altogether?
0.92	10	Truck A has a height of 4.5 metres and Truck B has a height of 4.3 metres. Which of the following shows the tunnels that both trucks can use safely?
0.95	14	Taylor desires to go on a hike that contains caverns and waterfalls and is longer than six kilometres. Which hike should Taylor undertake?

Goodness of fit

A Rasch Model analysis begins with determining the fitness of an item to the model. The result from Table 4 shows that 23 items fit the model, and 1 item did not fit the model. As for item number 3 that did not fit the model, we exclude it from further analysis as it has needed further revision. The other 23, although some items did not meet one of the requirements like item numbers 1, 2, and 7 that have outfit MNSQ greater than 1.3, they did meet the infit MNSQ, so it considered to fit the model.

Item Difficulty

According to Table 4, the item difficulty varies from -2.18 to 1.77 in the numerical reasoning examination. It demonstrates that 12 things, or 50% of the items with difficulty estimates with negative logits, are quite straightforward, but 12 items, or 50% of the items with difficulty estimates with positive logits, are relatively challenging. With a mean of.00 and a standard deviation of 1.23, the result demonstrates that there is a perfect correlation between simple and complex things. It also indicates that the difficulty limits are properly matched to the abilities of the students and that there is little variation in the students' scores. Other than item 14, which is the least difficult, the things are not arranged hierarchically according to their levels of difficulty. As requ ired by the Rasch Model for unidimensional scales, item 20, which is the most challenging item, should be the final item on the scale. As shown in the table 4, the item difficulty parameters are acceptable for the students' skills and there is little difference in the students' results.

Table 4
Item difficulty based on Rasch Result

Item	Item Measure (Difficulity)	Infit MNSQ	Outfit MNSQ	Goodness of Fit
1	-1.61	1.03	1.62	Fit
2	-.82	1.11	1.78	Fit
3	.58	1.40	1.52	Non Fit
4	-1.39	1.02	1.20	Fit
5	-.85	1.01	1.17	Fit
6	-1.17	1.04	1.11	Fit
7	-.60	1.03	1.62	Fit
8	-.94	.97	.90	Fit
9	-1.61	1.02	.86	Fit
10	-1.71	1.02	1.52	Fit
11	1.23	.93	.91	Fit
12	.54	.92	.85	Fit

13	-.02	.79	.67	Fit
14	-2.18	.95	1.10	Fit
15	1.47	.98	.98	Fit
16	-.37	.91	.73	Fit
17	.72	.86	.78	Fit
18	1.18	.92	.86	Fit
19	.05	.96	1.02	Fit
20	1.77	.92	.87	Fit
21	1.36	1.07	1.11	Fit
22	1.77	.91	.87	Fit
23	1.40	1.13	1.16	Fit
24	1.20	.89	.91	Fit
Mean	0.00			
SD	1.23			

The item difficulty parameters are not arranged hierarchically; item 14 should be the easiest item, followed by other things, and item 20 with a logit difficulty of 1.77 should be the final item on the scale for it to be a flawless unidimensional instrument.

Students' Ability

Measurements and time calculation

The number item for measurment and time calculation represesnts in number 5,8, 9, 13, 18, and 19. The question for item number five is to calculates elapsed time. This question used to test the ability of student in measurement and time calculation. About 84% of student is success in answering the question correctly. As the remaining 11% count the 7.40 to 8.15 as one hour and add it to another 15 minutes, the shallows analysis usually lead to this answer. The item number eight used to measure student ability in using measurement. About 85% student correctly answer this question and have the ability to uses measures and simple instrumen graduated in familiar units-ruler. The question for item number nine and thirdteen is related to performs a range of calculations to add and multiply lates elapsed time. In number nine About 92% of student is success in answering the question correctly. As the remaining 8% of students careless to add 2 fifty dollar and only calculate one dollar. While in item number thirdteen less than 80% of students answer the question correctly. About 10.4% answer 8100, which mean that they ignore the volume calcalution for ten minutes, 7.4% answer assume to find the volume after the crop watered, 3% of students were false in both calculate the time and interpret the question, the rest of them do not know the answer. Item number eightteen performs a range of calculations to multiply and add. Only 51% of the students correctly answer the questions. About 8.7% of students answered 1240.000 without add the extrapay for each hour and extrapay in the weekend, 24.7% answer 1360 without calculate the extra pay in the weekends, 4% answered 1940, and the rest of them about 10.8% did not know the answer. In number 19 is about convertion between metric units – kg to tonne. About 77% of them can answer this question correctly the rest of them 10.47%, 8.1 % careless about the diffrent unit in the question, 5% made mistakes in converting tons in kilograms and the rest of them did not know that the total means to add the variables not to substract the variables.

Mental Computation and Numerical Reason

The number item for mental computation and numerical reason represesnts in number 1, 4, 10,11,16,17,20. Item number one identifies and use whole number into thousands, this item represents the skill estimating and calculating using whole number. The students ability in answering this item is high, this proved by the percentage of students correclty answer this question which is about 91%, the remaining student choose c & d is nine students and b is one student. The mistake in this item mostly because of careless answer by the student. Item number four test the ability of student in using fraction, decimals, percentage, ratios and rates. The question is to interprets and uses familiar percentages. About 90% students are able to answer this questions correctly. While the remaining other doesn't know the concept of percentage. In item number ten almost all student (92%) can interpret the use of decimals, but only 51% can applies ratios which represent in number eleven. Both of number 16 and 17 asked student to uses fraction, about 69% of students can answer this question correctly. The rest of them make mistake in calculate the multiplication of a fraction.

Pattern and algebra

The number item for pattern and algebra in number 12 and 24. Item number twelve asks students to interprets and uses simple formulae that describe relationships between variables while item number twenty four spesificly ask students to interprets and uses simple formulae that describe relationships between variables. In these type of questions the result shows that less than 70% of students can correctly answer the question.

Spatial visualitation, geometric reasoning and mapping

The number item for spatial visualitation, geometric reasoning and mapping in number 2, 7, 15, 23. Item number two which is assesing the spatial visualitation, geometric reasoning and mapping by recognises feature of 2D shapes and 3D objects. Geometric concepts of all various types evolve over time, becoming increasingly integrated and synthesized as people learn to go beyond physical images and participate in 'taken-as-shared' mathematical discourse to describe, evaluate, infer, and deduce geometric relationships, possibly leading to formal proof (Siemon, Barkatsas, & Seah, 2019). For this question 84% of students answer correctly. Successful reasoners demonstrated a connected, integrated abstraction between numerical and geometric schemes, resulting in accomplishment in bothsurface and volume space reasoning (Seah & Horne, 2020). The other remaining 4% student didn't know which is the octagon shape, and about 9% of students counfuse it with the pentagon shapes, and the remaining 3% confuse it with rhombus and rectangle shapes. The Item number seven ask the student to read everyday maps its represent the ability in spatial visualitation, geometric reasoning and mapping. About 81% correctly answer this question, but 16% of students confuse it with anouter way as its also lead to desired destination but it needs one more direction. In item number 15 and 23, less that 50% of the students can answer correctly. Item number 15 related to geometry and 23 related to calculates and interprets information based on maps including scales.

Statistics and probability

The number item for statistics and probability in number 6, 14, and 22 The item number six ask students to identifies and compares information and data in tables For this item 88% students correctly answer the question, and the other 13% didn't pay attention to availibility of the hotel and take Blue Heaven (B) in answer as it provide both sail and surf but only in summer. The other 11% focused on the one without dots, and takes a big mistake as the one without dots mean it doesn't provide services needed by wayne. As the 6% that choose Diamond Shores (D) for an answer probably didn't pay much attention to the colour of the dots, and of course the dots play a big role as it's represent the availability. Item number fourteen Identifies and compares information and data in tables. Almost all the students (95%) can answer the question corrrectly. It means that the students found no difficulty in how to interpret the a data. Item number twentytwo Calculates and interprets mean in a variety of contexts. Most of the student (60%) could not calcultae and interpert mean in a data.

DISCUSSION

The purpose of the first study question was to evaluate the item difficulty characteristics of the numeracy tests in accordance with the Rasch Measurement Model. Although the result indicated that the item difficulty parameters were suitable for measuring the students' skills, they were not ordered hierarchically from the least difficult to the most difficult item, as required by the Rasch Measurement Model for unidimensional scales. As required by the Rasch Model for unidimensional tests, the students were not given the chance to answer from easier to more difficult questions, which might have resulted in a loss of interest and motivation on the part of the students.

This conclusion is not supported by the findings of Phillipson (2008), as she discovered in her research that the most difficult items that comprised the scales for her research were located at the lower end of the scales, while the easiest items were located at the beginning, as required by the Rasch Model for undimensional scales. This finding contradicts the findings of HermanAbell and DeBoer (2011), who reported in their research that the items comprising the instrument were arranged according to their increased complexity, albeit a small number of items exhibited redundancy, and who concluded that the instrument was unidimensional as a whole.

The purpose of the second study topic was to investigate the students' ability parameters on the Numeracy test using the Rasch Measurement Model. As indicated by the fact that the students' logits fell within the range advised by Rasch analysis and by the fact that the students were ranked from least capable to most capable based on their test scores, the results indicate that the students' abilities were accurately judged. When an examination's components assess a unidimensional construct, it is easiest to compare the talents of different students. This conclusion is consistent with that of Nopiahet al. (2010), who determined in their research that the items comprising tests accurately measured students' skills. This result is also consistent with the findings of Khairaniet al. (2012), who found that the students' ability estimations were consistent with the Rasch Measurement Model's expectations and that the items contributed meaningfully to evaluating the students' abilities.

The most challangging ability for students were calculates and interprets mean in variaty of context and perform a range of calculation to add, subctract and multiply especially in performs a range of calculations to add, subtract and multiply and calculates and interprets mean in a variety of contexts.

The students have high performance in identifies and compares information and data in tables.

RELEVANCE

The Rasch Measurement Model reduces the score categories to two binary groups. This boosts the accuracy of the model for the examined things and individuals. The implication of this work is that item and person parameter estimations and other psychometric traits will be measured with great precision. Using these data, the performance, strengths, and shortcomings of students on exams might be effectively explained. Examining bodies and test analysts can utilize the Rasch Measurement Model's many psychometric features, such as item difficulty and individual ability factors.

The Rasch Model creates valid and reliable examination items with suitable item difficulty and person ability parameters that assess the target unidimensional construct.

Since the research enables the identification of each examinee's testing strengths and weaknesses, this diagnostic should be used to enhance the quality of students' numeracy. Two significant techniques are offered for enhancing the quality of numeracy. Teaching treatment: Mathematics teachers should tailor their instructional materials to the ability patterns of their students. For example, if a student has limited knowledge of algebra or numbers topic, a presentation stressing the use of diagrams, models, and actual examples is recommended.

Instruction with feedback and correction procedures: Feedback devices, such as formative evaluation and diagnostic exams, should be incorporated into the instruction to discover weaknesses in the students' understanding of a particular unit of numeric. Reteaching a specific unit of mathematics knowledge, small-group study sessions, and individual tutoring are the most effective remedial strategies.

CONCLUSION AND IMPLICATION

The Rasch Model gives a comprehensive analysis of test item and individual factors. Our findings provide guidance for individuals at University as well as other schools who provide exam preparation assistance to students. Specifically, supports should focus on updating test-taking abilities, particularly in numeray test experience. In addition, boosting mathematics abilities, establishing ways to comprehend numeracy settings and questions, and enhancing computation fluency will help students excel not only on the numeracy test, but also in their future jobs and in everyday life. Such aids may be especially essential for students who have not engaged in a substantial amount of mathematics study at the higher levels of secondary school or university, or in recent years. This group may comprise elementary preservice teachers and secondary preservice teachers whose specializations do not emphasize mathematical thinking. Students should graduate with greater competence, self-assurance, and critical awareness of mathematics in their curriculum and career.

Disclosure statement

No potential conflict of interest was reported by the authors

ACKNOWLEDGMENTS

I am grateful to all of those with whom I have had the pleasure to work during this projects. I would like to express my deepest appreciation to Direktorat Penelitian dan Pengabdian Masyarakat Universitas Muhammadiyah Gresik for funding this project.

REFERENCES

Bråting, K., & Kilhamn, C. (2021). Exploring the intersection of algebraic and computational thinking. *Mathematical Thinking and Learning*, *23*(2), 170–185. https://doi.org/10.1080/10986065.2020.1779012

Carbonell Carrera, C., & Bermejo Asensio, L. A. (2017). Landscape interpretation with augmented reality and maps to improve spatial orientation skill. *Journal of Geography in Higher Education*, *41*(1), 119–133. https://doi.org/10.1080/03098265.2016.1260530

De-Winter, J. C. F., & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics*, *39*(4), 695–710. https://doi.org/10.1080/02664763.2011.610445

Erdem, E., & Gürbüz, R. (2014). An Analysis Of Seventh-Grade Students' Mathematical Reasoning. *Journal*, *44*(1), 123–142.

Fagerlin, A., Zikmund-fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring Numeracy without a Math Test : Development of the Subjective Numeracy Scale. *Medical Decision Making*, 672–680. https://doi.org/10.1177/0272989X07304449

Gal, I. (2005). Towards "Probability Literacy" for all Citizens: Building Blocks and Instructional Dilemmas. In G. A. Jones (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning* (pp. 39–63). Boston, MA: Springer US. https://doi.org/10.1007/0-387-24530-8_3

Galligan, L. (2013). A systematic approach to embedding academic numeracy at university. *Higher Education Research and Development*, *32*(5), 734–747. https://doi.org/10.1080/07294360.2013.777037

Gregory, L., Villarosa, A. R., Ramjan, L. M., Hughes, M., O'Reilly, R., Stunden, A., Daly, M., Raymond, D., Fatayer, M., & Salamonson, Y. (2019). The influence of mathematics self-efficacy on numeracy performance in first-year nursing students: A quasi-experimental study. *Journal of Clinical Nursing*, *28*(19–20), 3651–3659. https://doi.org/https://doi.org/10.1111/jocn.14963

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, *32*(2), 175–191. https://doi.org/https://doi.org/10.1016/j.intell.2003.12.001

Henson, R. K., Capraro, R. M., & Capraro, M. M. (2004). Reporting practices and use of exploratory factor analyses in educational research journals: Errors and explanation. *Research in the Schools*, *11*(2), 61–72.

Jarvis, C. H. (2011). Spatial Literacy and the Postgraduate GIS Curriculum. *Procedia - Social and Behavioral Sciences*, *21*, 294–299. https://doi.org/https://doi.org/10.1016/j.sbspro.2011.07.022

Jonas, N. (2018). Numeracy Practices and Numeracy Skills among Adults. *OECD Education Working Paper No. 177*, (177). Retrieved from http://dx.doi.org/10.1787/8f19fc9f-en

Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences*, *20*(2), 82–88. https://doi.org/10.1016/j.lindif.2009.07.004

Kaput, J. (2008). What is algebra? What is algebraic reasoning? In J. Kaput, D. Carraher, & M. Blanton (Eds.), *Algebra in the early grades* (pp. 5–18). Lawrence Erlbaum.

Kieffer, K. M. (1999). An introductory primer on the appropriate use of exploratory and confirmatory factor analysis. *Research in the Schools*, *6*(2), 75–92.

Lee, L., & Freiman, V. (2006). Developing Algebraic Thinking through Pattern Exploration. *Mathematics Teaching in the Middle School*, *11*(9), 428–433. https://doi.org/10.5951/MTMS.11.9.0428

McCowan, R. J., & McCowan, S. C. (1999). *Item Analysis for Criterion-Referenced Tests*. New York: Center Development of Human Services.

Mohamad, M. M., Lisa, N., Sern, L. C., & Mohd, K. (2015). Measuring the Validity and Reliability of Research Instruments. *Procedia - Social and Behavioral Sciences*, *204*(November 2014), 164–171. https://doi.org/10.1016/j.sbspro.2015.08.129

N.A. (2019). Learn About Reliability Analysis and Factor Analysis in SPSS With Online Shopping Data (2006). *SAGE Research Methods Datasets*. https://doi.org/10.4135/9781526468321

Newcombe, N. S., & Shipley, T. F. (2015). Thinking About Spatial Thinking: New Typology, New Assessments BT - Studying Visual and Spatial Reasoning for Design Creativity. In J. S. Gero (Ed.) (pp. 179–192). Dordrecht: Springer Netherlands.

OECD. (2015). *Skill Matters: Furthe Results From the Survey of Adults Skills*. Jakarta. Retrieved from https://www.oecd.org/skills/piaac/Skills-Matter-Jakarta-Indonesia.pdf

OECD. (2016). *Skill Matter: Further Results from the Survey of Adult Skills*. Paris. Retrieved from http://dx.doi.org/10.1787/9789264258051-en

Prince, R., & Frith, V. (2020). An investigation of the relationship between academic numeracy of university students in South Africa and their mathematical and language ability. *ZDM*, *52*, 433–455. https://doi.org/10.1007/s11858-019-01063-7

Rothman, R. L., Housam, R., Weiss, H., Davis, D., Gregory, R., Gebretsadik, T., … Elasy, T. A. (2006). Patient Understanding of Food Labels: The Role of Literacy and Numeracy. *American Journal of Preventive Medicine*, *31*(5), 391–398. https://doi.org/https://doi.org/10.1016/j.amepre.2006.07.025

Samuels, P. (2016). *Advice on Exploratory Factor Analysis. Centre for Academic Success, Birmingham City University*. Retrieved from https://www.researchgate.net/profile/Peter_Samuels/publication/319165677_Advice_on_Exploratory_Factor_Analysis/links/5996c447a6fdcc35c6c954dc/Advice-on-Exploratory-Factor-Analysis.pdf?origin=publication_detail

Scholz, M. A., Huynh, N. T., Brysch, C. P., & Scholz, R. W. (2014). An Evaluation of University World Geography Textbook Questions for Components of Spatial Thinking. *Journal of Geography*, *113*(5), 208–219. https://doi.org/10.1080/00221341.2013.872692

Seah, R. T. K., & Horne, M. (2020). The influence of spatial reasoning on analysing about measurement situations. *Mathematics Education Research Journal*, *32*(2), 365–386. https://doi.org/10.1007/s13394-020-00327-w

Siemon, D., Barkatsas, T., & Seah, R. (2019). *Researching and Using Progressions (Trajectories) in Mathematics Education*. Leiden, The Netherlands: Brill. https://doi.org/https://doi.org/10.1163/9789004396449

Singh, P., Hoon, T. S., Nasir, N. A. M., Han, C. T., Rasid, N. S. M., & BZH, J. (2020). An Analysis of Students' Mathematical Reasoning and Mental Computation Proficiencies. *Universal Journal of Educational Research*, *8*(11), 5628–5636. https://doi.org/10.13189/ujer.2020.081167

Sunderaraman, P., Barker, M., Chapman, S., & Cosentino, S. (2022). Assessing numerical reasoning provides insight into financial literacy. *Applied Neuropsychology: Adult*, *29*(4), 710–717. https://doi.org/10.1080/23279095.2020.1805745

Vignoles, A. (2016). What is the economic value of literacy and numeracy? *IZA World of Labor*, (229), 1–10. https://doi.org/10.15185/izawol.229

Wetzel, A. P. (2011). *Factor Analysis Methods And Validity Evidence: A Systematic Review Of Instrument Development Across The Continuum Of Medical Education*. Virginia Commonwealth University. Retrieved from http://scholarscompass.vcu.edu/etd/2385

Windisch, H. C. (2015). Adults with low literacy and numeracy skills: A literature review on policy interventions. *OECD Education Working Papers No. 123*, *12*(123), 2–125. https://doi.org/https://dx.doi.org/10.1787/5jrxnjdd3r5k-en

Windisch, H. C. (2016). How to motivate adults with low literacy and numeracy skills to engage and persist in learning: A literature review of policy interventions. *International Review of Education*, *62*(3), 279–297. https://doi.org/10.1007/s11159-016-9553-x