# Items Analysis on Mathematics Online Final Examination Using the Rasch Measurement Model in an Indonesian Junior High School

**Tommy Tanu Wijaya**
*Beijing Normal University, Beijing,* China
*E-mail : 202139130001@mail.bnu.edu.cn*

**Jiajie Yan**
*Beijing Normal University, Beijing,* China
*E-mail : jjyan@mail.bnu.edu.cn*

**In In Mutmainah***
*SMP N 33 Bandung,* Indonesia
*E-mail : in.mutmainah@gmail.com*

**M. Afrilianto**
*Institut Keguruan dan Ilmu Pendidikan Siliwangi, Cimahi,* Indonesia
*E-mail : muhammadafrilianto1@gmail.com*

**Tina Rosyana**
*Institut Keguruan dan Ilmu Pendidikan Siliwangi, Cimahi,* Indonesia
*E-mail : tinarosyana@gmail.com*

**Xinxin Li**
*Guangxi Normal University, Guilin,* China
*E-mail : 1748308781@qq.com*


*DOI :* http://dx.doi.org/10.30983/educative.v7i2.6001
**Corespondent Auhtor*

| Submission : November 11, 2022 | Revished : November 28, 2022 | Accepted : December 24, 2022 | Published : December 30, 2022 |
|---|---|---|---|

## Abstract

The National Examination in Indonesia has been abolished since 2020. Indonesian Junior High Schools (SMP) now hold its own final examination items for the 9th-grade, and from the results, the school determines the graduation an success of students. The final examination has an important role and significant value in making decisions about students' graduation. This study analyzes the Online Final Examination items in one of the public junior high schools in the city of Bandung, West Java. The sample was 234 students in grade 9 in regard to their mathematics examination tests, which comprises of 20 multiple-choice items with four options (A-D). Data processing was carried out using Winsteps software with the Rasch modeling technique. Subsequently, the Rasch model results showed an acceptable students separation statistic of 1.54 and a poor person reliability of 0.7<0.8. The item separation statistics was in a good category at 4.59, while reliability was at 0.86, which was excellent. Although four online final examination items were in the fit category, 16 were good and capable of dividing students according to their abilities. They also provided very detailed data about the quality of the items and the ability of each grade 9 student. Since each test item is in the fit category, this study offers advice on how teachers should prepare and analyze the Online Final Examination.

**Keywords**: Final Examination, Junior High School, Mathematics Course, Rasch Model

## *Abstrak*

*Ujian nasional di indonesia sudah ditiadakan sejak tahun 2020. Satuan pendidikan tingkat SMP selanjutnya menentukan sendiri ujian akhir untuk siswa kelas 9 dan menentukan kelulusan mereka. Ujian akhir adalah penentuan dan tahap penting untuk menentukan kelulusan siswa. study ini bertujuan untuk menganalisis ujian akhir matematika secara online di salah satu SMP di Bandung, jawa barat. Sample pada penelitian ini adalah 234 siswa kelas 9. Ujian matematika berisi 20 soal pilihan ganda dengan 4 options. Data yang berhasil dikumpulkan diolah dengan menggunakan software winsteps dengan Teknik Rasch model. hasil penelitian menunjukkan bahwa students separation statistic adalah 1.54 dan poor person reliability of 0.7<0.8. item separation statistics termasuk*

*pada kategori baik dengan nilai 4.59, dimana mempunyai reabilitas sangat baik dengan nilai 0.86. meskipun kami menemukan 4 soal ujian pada kategori cukup, tetapi 16 soal ujian termasuk pada kategori baik untuk mengetes kemampuan matematika siswa. penelitian ini juga menyediakan analisis data dengan detail untuk setiap butir soal ujian dan kemampuan setiap siswa. penelitian ini dapat berkontribusi dengan memberi contoh cara menganalisis soal ujian akhir..*

***Kata Kunci**: Ujian Akhir, SMP, Pembelajaran Matematik, Rasch Model*

## Introduction

Indonesia is one of the countries in Southeast Asia with a high number of Covid-19 cases, and as of May 18, 2020, there were 1.7 million cases and 48,305 deaths. This has caused many challenges and problems in education, as students have not attended school since May 2020, and teaching and learning activities have taken place online [1]. Several problems have arose when implementing online learning, as the teachers did not have good pedagogical and technological knowledge to teach on these platforms [2]. Also, students were not ready to take online lessons [3].

National examinations for junior high school are usually held in many countries to determine students' graduation and future when entering senior high school or university [4]. Similar to other countries, Indonesian student's national examination was held every year in to determine the students' graduation and analyze their quality. The Indonesian Ministry of Education decided to abolish the national examinations from elementary to high school level in 2021 [5]. Therefore, each school is allowed to hold a final examination and determine the students' graduation independently. This development can be beneficial depending on the angle viewed by experts and teachers. Consequently, the government cannot determine the quality of education in each region anymore because every school has different graduation examinations and standards. The negative aspect of the policy is that schools may assess a child's graduation subjectively and pass a child with poor achievement scores. Since this problem is yet to be solved, analyzing and researching the quality of final examination items for grade 9 is an important endeavor to undertake.

The challenges of final graduation examination are even greater because they are held online, using Google Classroom, Zoom, Google Meeting, and other applications [6]. Furthermore, the examination is usually in multiple-choice and description form inputted into the Google platform to allow students to work online. In the end, the teacher does not have to check the multiple-choice answers, as students can immediately see the final results.

[1] Purniadi Putra, Fahrina Yustiasari Liriwati, and Tasdin Tahrim, 'The Students Learning from Home Experience during Covid-19 School Closures Policy in Indonesia', *Jurnal Iqra': Kajian Ilmu Pendidikan*, 5.2 (2020), 30–42; Zetra Hainul Putra, Gustimal Witri, and Intan Kartika Sari, 'Prospective Elementary Teachers' Perspectives on Online Mathematics Learning during Coronavirus Outbreak', *Journal of Physics: Conference Series*, 1655.1 (2020) <https://doi.org/10.1088/1742-6596/1655/1/012057>.

[2] T.T. Wijaya, Zhou Ying, and Lin Suan, 'Gender and Self-Regulated Learning During COVID-19 Pandemic in Indonesia', *Jurnal Basicedu*, 4.3 (2020), 725–32 <https://doi.org/10.31004/basicedu.v4i3.422>.

[3] T.T. Wijaya and others, 'Indonesian Students' Learning Attitude towards Online Learning during the Coronavirus Pandemic', *Psychology, Evaluation, and Technology in Educational Research*, 3.1 (2020), 17–25 <https://doi.org/10.33292/petier.v3i1.56>.

[4] Xiaohong Wu, Ying Zhou, and Zongzhao Mo, 'A Comparative Study on the Comprehensive Difficulty of Junior High School National Examination', *Journal On Education*, 2.4 (2020), 352–66; Sardjana Orba Manullang and others, 'The Review of the International Voices on the Responses of the Worldwide School Closures Policy Searching during Covid-19 Pandemic', 5.2 (2020), 1–13.

[5] thomas harming Suwarta, 'Ini Sejarah Ujian Nasional Di Indonesia', *Media Indonesia* (jakarta, 2019).

[6] Mengzhou Li and others, 'Optimized Collusion Prevention for Online Exams during Social Distancing', *Npj Science of Learning*, 6.1 (2021) <https://doi.org/10.1038/s41539-020-00083-3>; Digvijay Pandey and others, 'COVID-19: A Framework for Effective Delivering of Online Classes During Lockdown', *Human Arenas*, 0123456789, 2021 <https://doi.org/10.1007/s42087-020-00175-x>.

However, Indonesia does not have a stable internet network and an effective platform to ensure students do not cheat during the examination [7]. They may use calculators while solving the math questions or cellphones to ask friends for answers, and even people around them can help them get perfect scores. Furthermore, the analysis of these results is very important to evaluate the students' abilities and ensure cheating does not occur.

This research focuses on analyzing the use of online mathematics examination to replace the national examinations in determining student graduation. The novelty of the research consists in the analysis of two important points: the items' quality and students' abilities using Rasch analysis with mini-step software. Furthermore, the research was conducted on a public school in Bandung, Indonesia, using the Rasch measurement model to evaluate the difficulty, reliability, and quality of items. It analyzes the probability curve and students' abilities. Consequently, the results are expected to contribute to the field of education since it evaluates whether the abolishment of national examination has been successful . They can also serve as a reminder for teachers when preparing items for final examinations, such as midterms, semester finals, or online mathematics finals.

By processing the final examination data using the Rasch measurement model, the results can answer the following questions:

1. What is the student achievement in the mathematics online final examination?
2. What are the quality and difficulty level of online final mathematics examination items?
3. Can the online final mathematics examination items determine the students' ability levels?
4. Are there any indications of cheating when students work on the mathematics online final examination?

## Research Theory

### *National Examination (UN)*

Every year, many countries use the national examination to determine the graduation of grade 9 students in junior or high schools. These countries have their designations for this examination. In the United States, for instance, thisis called the Scholastic Aptitude Tests (SAT) or American College Testing (ACT) [8]. In China, it is known as the Chinese National College Entrance Examination (CEE; *gaokao*) [9] and in Indonesia, it is called the *Ujian Nasional* (UN) [10].

The national examinations are conducted once yearly around May and are important for most students to determine their future and further their education. In Indonesia, students who fail can retake it in the following year. The national examination is used to evaluate students, schools, and provinces. It helps the government assess the quality of education. From the results, the government can focus on areas that require special attention and improvement to ensure that education in the country is fair and equitable.

Several studies on national examinations have been published, such as LoCastro [11] that compare the national examination on the English subject in Japan using sociocultural analysis. Another study, Bai et al. [12] examined the National College Entrance

[7] Wijaya and others; Dwi Prasetyanto, Muhamad Rizki, and Yos Sunitiyoso, 'Online Learning Participation Intention after COVID-19 Pandemic in Indonesia: Do Students Still Make Trips for Online Class?', *Sustainability (Switzerland)*, 14.4 (2022) <https://doi.org/10.3390/su14041982>.

[8] H. H. Hohne, 'The Prediction of Academic Success', *Australian Journal of Psychology*, 1.1 (1949), 38–42 <https://doi.org/10.1080/00049534908256014>.

[9] Wu, Zhou, and Mo.

[10] Agus Purwanto and others, 'Studi Eksploratif Dampak Pandemi COVID-19 Terhadap Proses Pembelajaran Online Di Sekolah Dasar', *EduPsyCouns Journal*, 2.1 (2020).

[11] Virginia Locastro, 'The English in Japanese University Entrance Examinations: A Sociocultural Analysis', *World Englishes*, 9.3 (1990), 343–54 <https://doi.org/10.1111/j.1467-971X.1990.tb00271.x>.

[12] Chong-en Bai, Wei Chi, and Xiaoye Qian, 'China Economic Review Do College Entrance Examination Scores Predict Undergraduate GPAs ? A Tale of Two Universities',

Examination and analyzed students' abilities at two leading universities in China, with implications for university admission policies and practices. This study suggested that admissions should consider the National College Entrance Examination results, alongside the high school achievement over the past three years, take into account the prizes and awards received in high school, and consider homeroom teacher recommendations. Although the National College Entrance Examination score is just a number, the many factors it includes should be reconsidered, for example, general and language skills and attitudes.

Sanz [13] has shown that the national foreign language examination system in Spain uses an online system. The illustration of the platform, management tools, security, and user interaction are well explained, and the study observed that the use of online systems for national examinations has existed before the coronavirus pandemic.

Meanwhile, no other research has been published that deals with the item analysis of an online national examination u. This research deals with an online national examination item analysis during Covid-19. The analysis can also be used as a reference source, benchmark, and comparison material to assess the quality of national examination items in Indonesia during Covid-19. a public school was used as the research sample.

### *The Rasch Model*

The Rasch model is a mathematical model and measurement tool [14]. It focuses on the approach to construct measurement in the social sciences, which usually uses the more familiar mini-step software. It can show the response structure of the assessment [1516] Moreover, it can enter the class, student work, questionnaire items, or final examinations and provide specific measurement data [17]. Although the Rasch model shows the measurement criteria and test fitting responses, the analysis of insufficiently fit data should be continued by carefully evaluating reliability and validity [18]. The model is also a method for analyzing examination results to investigate the correlation between item difficulty and students' abilities [19].

Meanwhile, Omar and Sayaka Karlin [20] stated the importance of validating test items. They affirmed that the validity test can determine the accuracy of test items in measuring students'

*China Economic Review*, 30 (2014), 632–47 <https://doi.org/10.1016/j.chieco.2013.08.005>.

[13] Ana Gimeno Sanz and Ana Sevilla Pavón, 'Toward Implementing Computer-Assisted Foreign Language Assessment in the Official Spanish University Entrance Examination', in *Proceedings of the 2015 EUROCALL Conference, Padova, Italy*, 2015, pp. 215–20.

[14] N. Hermita and others, 'Level Conceptual Change Pre-Service Elementary Teachers on Electric Current Conceptions through Visual Multimedia Supported Conceptual Change', *Journal of Physics: Conference Series*, 1013.1 (2018), 1–8 <https://doi.org/10.1088/1742-6596/1013/1/012060>; Ching Sing Chai, Xingwei Wang, and Chang Xu, 'An Extended Theory of Planned Behavior for the Modelling of Chinese Secondary School Students' Intention to Learn Artificial Intelligence', *Mathematics*, 8.11 (2020), 1–18 <https://doi.org/10.3390/math8112089>.

[15] Joshua A. Weller and others, 'Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach', *Journal of Behavioral Decision Making*, 26.2 (2013), 198–212 <https://doi.org/10.1002/bdm.1751>.

[16] .

[17] Siti Eshah Mokshein, Haliza Ishak, and Hishamuddin Ahmad, 'The Use of Rasch Measurement Model in English Testing', *Cakrawala Pendidikan*, 38.1 (2019), 16–32 <https://doi.org/10.21831/cp.v38i1.22750>.

[18] Shun Chin Yang and others, 'Statistical Item Analysis of the Examination in Anesthesiology for Medical Students Using the Rasch Model', *Journal of the Chinese Medical Association*, 74.3 (2011), 125–29 <https://doi.org/10.1016/j.jcma.2011.01.027>.

[19] Ayu Faradillah and Leha Febriani, 'Mathematical Trauma Students' Junior High School Based on Grade and Gender', *Infinity Journal*, 10.1 (2021), 53 <https://doi.org/10.22460/infinity.v10i1.p53-68>; Haliza Othman and others, 'Application of Rasch Measurement Model in Reliability and Quality Evaluation of Examination Paper for Engineering Mathematics Courses', *Procedia - Social and Behavioral Sciences*, 60.2009 (2012), 163–71 <https://doi.org/10.1016/j.sbspro.2012.09.363>.

[20] Omar Karlin and Sayaka Karlin, 'Making Better Tests with the Rasch Measurement Model', *InSight: A Journal of Scholarly Teaching*, 13.1 (2018), 76–100 <https://doi.org/10.46504/14201805ka>.

abilities, for instance, in mathematics. Validation can also ascertain whether the difficulty level of the test items is based on the student's abilities. Can the test items separate the students into 3 or 4 levels based on their abilities? Are the test items self-explanatory? Do test items confuse students? Their research concluded that ensuring test items given are appropriate and effective in measuring students' abilities is important.

Subsequently, the number of items with high, medium, and low difficulty should be planned properly, in order that the student's ability level can be measured specifically [21]. The test items that are too difficult will achieve mostly incorrect answers, while students will correctly answer questions that are too easy[22]. These conditions cannot measure which students have low, medium, or very good abilities. Therefore, test items should be created with 50% medium difficulty, 30% easy, and 20 percent high difficulty.

The item analysis using the Rasch model has been performed in many previous studies in various fields such as medicine [23], pharmacy [24], physics [25], chemistry [26], social science, and so on to validate and evaluate the quality of the questionnaire items. However, only a few studies have used the Rasch model to analyze items in the mathematics field, while none that analyzed final school examination items using the model was discovered [27].

A study on the use of the Rasch model to validate and analyze items in 2011 was found. Mohsen Tavakol & Reg Dennick used the model to improve assessment in medical education by analyzing 355 medical students using 24 final clinical knowledge items [28]. The analysis results concluded that Rasch analysis supports the diagnosis of quality. It also provided feedback for each test item and students' ability to inform lecturers on methods to improve the quality of examination items. In 2011 again, Huda Abdullah et al. analyzed examination items on Microelectronic material using the Rasch model and concluded it is capable of analyzing students' abilities when answering examination items [29].

---

[21] H. Susanto, A Rinaldi, and Novalia, 'Analisis Validitas Reabilitas Tingkat Kesukaran Dan Daya Beda Pada Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika', *The Journal of the Japan Society for Respiratory Endoscopy*, 6.2 (2015) <https://doi.org/10.18907/jjsre.37.3_343_4>; Weller and others.

[22] Bagiyono, 'Analisis Tingkat Kesukaran Dan Daya Pembeda Sial Ujian Pelatihan Radiografi Tingkat 1', *Widyanuklida*, 16.No. 1 (2017), 1–12.

[23] Dedy Triono, Riyanarto Sarno, and Kelly R. Sungkono, 'Item Analysis for Examination Test in the Postgraduate Student's Selection with Classical Test Theory and Rasch Measurement Model', *Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, ISemantic 2020*, 2020, 523–29 <https://doi.org/10.1109/iSemantic50169.2020.9234204>.

[24] Jean Benoit Hardouin, 'Rasch Analysis: Estimation and Tests with Raschtest', *Stata Journal*, 7.1 (2007), 22–44 <https://doi.org/10.1177/1536867x0700700102>.

[25] Lijuan Gan and Tommy Tanu Wijaya, 'MEASURING STUDENT'S LOGICAL REASONING SKILLS OF CHINESE SENIOR HIGH SCHOOL USING RASCH MEASUREMENT MODEL', *JURNAL MathEdu (Mathematic Education Journal)*, 4.3 (2021), 142–49.

[26] Rahmi Qurota Aini, Arif Rachmatullah, and Minsu Ha, 'Indonesian Primary School and Middle School Students' Attitudes toward Science: Focus on Gender and Academic Level', *Journal of Baltic Science Education*, 18.5 (2019), 654–67 <https://doi.org/10.33225/jbse/19.18.654>.

[27] Hasni Shamsuddin and Ahmad Zamri Khairani, *Proceedings of the Regional Conference on Science, Technology and Social Sciences (RCSTSS 2016)*, *Proceedings of the Regional Conference on Science, Technology and Social Sciences (RCSTSS 2016)* (Springer Singapore, 2019) <https://doi.org/10.1007/978-981-13-0203-9>; Liu Huang and others, 'Constructs Evaluation of Student Attitudes towards Science', *Eurasia Journal of Mathematics, Science and Technology Education*, 15.12 (2019) <https://doi.org/10.29333/ejmste/109168>; Richard L. Haspel and others, 'Development of a Validated Exam to Assess Physician Transfusion Medicine Knowledge', *Transfusion*, 54.5 (2014), 1225–30 <https://doi.org/10.1111/trf.12425>; Heri Retnawati and Nidya F. Wulandari, 'The Development of Students' Mathematical Literacy Proficiency', *Problems of Education in the 21st Century*, 77.4 (2019), 502–14 <https://doi.org/10.33225/pec/19.77.502>.

[28] Mohsen Tavakol and Reg Dennick, 'Psychometric Evaluation of a Knowledge Based Examination Using Rasch Analysis: An Illustrative Guide: AMEE Guide No. 72', *Medical Teacher*, 35.1 (2013), e838–48 <https://doi.org/10.3109/0142159X.2012.737488>.

[29] Tavakol and Dennick.

Also, Nicholas and team in 2016, analyzed multiple-choice examination items conducted on 101 students at Arizona University and found that they were not very difficult, hence the classification of students' abilities was not measured properly [30]. They advised that the item-making should focus on measuring the performance quality rather than just looking at student grades and rankings. Furthermore, research by Haliza Othman on Engineering Mathematics Courses on the code paper KKKQ2114 used the model to validate and analyze items' quality and found that 10% were in the misfit category [31]. Therefore, the problem should be corrected or removed, as the bad items cannot measure students' abilities according to the teacher's wishes.

Based on these previous studies, this research concluded that the Rasch model is very important and has many benefits for evaluating and improving the quality of items as well as measuring and analyzing students' abilities. Furthermore, the results from the model can be used as evaluation material to modify the teaching method or the item forms to fit the guidelines for good test items.

## Research Method

This study uses the Rasch model to analyze test items from the Online Final Examination data. The measurement model can show summary statistics such as mean, Standard Deviation, maximum and drinking value, amount of data, etc., and classify students' abilities as low, medium, or high by evaluating separation and logit. It can also determine whether the items for obtaining a measurement were met, the level of difficulty, and whether students are careful when answering or making guesses.

### *Research Population*

The research data comprises of mathematics examination items for the 9th-grade graduation class held in May 2021. Meanwhile, the sample was the 9th-grade students in one of the public junior high schools in Bandung, Indonesia. The purposive sampling technique was used in the selection, and the general information about the study sample is shown in Table 1. This public school was chosen because it has good national accreditation and quality.

**Table 1**. General Information

| Category | Public School |
|---|---|
| Accreditation | A |
| Number of students | 930 students |
| Status | National standard school |
| Lowest National Examination Score (2015) | 270/400 |
| Location | Bandung, West Java, Indonesia |

The final examination items were made by the 9th-grade math teachers and validated by 2 curriculum experts, then signed by the principal and used to determine the students' graduation in mathematics. Subsequently, the research sample was 234 students from six classes (9A, 9B, 9C, 9D, 9E, 9F), comprising 155 female and 79 male students, which makes up 66.24% and 33.76%, respectively.

## Data collection

A teacher with the initials IM made 20 graduation examination items for mathematics in the multiple-choice form, consisting of 6 items on numbers, algebra, and geometry each, and 2 items on probability and statistics each. Table 2 shows more specific information.

[30] Nicholas B. Jennings and others, 'Measurement Characteristics of a Concept Classification Exam Using Multiple Case Examples: A Rasch Analysis', *Currents in Pharmacy Teaching and Learning*, 8.1 (2016), 31–38 <https://doi.org/10.1016/j.cptl.2015.09.010>.

[31] Zulkifli Mohd Nopiah and others, 'Identification of Student Achievement and Academic Profile in the Linear Algebra Course: An Analysis Using the Rasch Model', *2011 3rd International Congress on Engineering Education: Rethinking Engineering Education, The Way Forward, ICEED 2011*, 2011, 197–202 <https://doi.org/10.1109/ICEED.2011.6235389>.

**Table 2.** The final examination material according to Indonesian national standards

| Nbr. | Tested competence | Scope of material | Material | Cognitive level |
|---|---|---|---|---|
| 1 | The students can understand and are knowledgeable about integer operations | Numbers | Integer operations | Knowledge and understanding |
| 2 | The students can understand and are knowledgeable about quadratic operations | Numbers | Quadratic operation | Knowledge and understanding |
| 3 | The students can apply their knowledge of fractional numbers | Numbers | Fractional number | Application |
| 4 | The students can apply their knowledge of comparisons | Numbers | Comparison | Application |
| 5 | The students can apply their knowledge of social arithmetic | Numbers | Social arithmetic | Application |
| 6 | The students can apply their knowledge of number sequences and series | Numbers | Sequences and series of numbers | Reasoning |
| 7 | The students can understand and are knowledgeable about the linear inequality of one variable | Algebra | Linear inequality of one variable | Knowledge and understanding |
| 8 | The students can understand and are knowledgeable about sets of numbers | Algebra | Set of numbers | Knowledge and understanding |
| 9 | The students can use reasoning related to straight-line equations | Algebra | Quadratic function | Knowledge and understanding |
| 10 | The students can apply algebraic forms | Algebra | Algebra forms | Application |
| 11 | The students can apply their knowledge of algebraic forms | Algebra | Quadratic function | Application |
| 12 | The students can apply their knowledge of relations or functions | Algebra | Function value | Application |
| 13 | The students can apply their knowledge of geometry and measurement | Geometry and measurement | Pythagorean Theorem | Application |
| 14 | The students can understand and are knowledgeable about lines and angles | Geometry and measurement | Lines and angles | Knowledge and understanding |
| 15 | The students can understand and are knowledgeable about triangles' similarity and congruence | Geometry and measurement | Triangles similarities and congruence | Knowledge and understanding |
| 16 | The students can apply their knowledge of curved side spaces | Geometry and measurement | Curved side spaces | Application |
| 17 | The students can apply their knowledge of circles | Geometry and measurement | Circles | Application |
| 18 | The students can understand transformation knowledge | Geometry and measurement | Transformation | Knowledge and understanding |
| 19 | The students can understand the data presentation of frequency table forms | Statistics and probability | Data centering measure | Knowledge and understanding |
| 20 | The students can understand the probability of events | Statistics and probability | Probability of events | Application |

Each test item was assigned a 5-point score, and the student's total will range from 5 to 100. Based on the results, this study used the Rasch measurement model to classify the students' achievement when working on the Online Final Examination items.

## Result and Disussions

### Process

This research analyzed the teaching and learning process by interviewing two mathematics teachers at the school. According to the results, these activities were still occurring online, and the school used Google Meeting and Google Classroom, and occasionally Zoom to interact with students. Meanwhile, homework and exercises were given through Google Classroom. From the interview results, the teachers stated that no math

software or videos were used for mathematics lessons during the Covid-19 pandemic.

The 9th-grade final examination information was obtained through Google Meeting and lasted for 90 minutes, with students using laptops to work on the items and a handphone placed behind them for supervision. However, not all the students have these gadgets, and some only used handphones to take the final examination. Another problem was unreliable internet connection, which prevented the teacher from monitor the students properly. Finally, the instructions were given to students to log in at 9 P.M and the test was concluded at 10.30 P.M.

### Statistical analysis

This research used the Person fit statistics to evaluate the consistency of the students' answers in the online examination. A correct answer means the test item is equal to or below the student's ability, while an incorrect answer means higher. Therefore, an inconsistent Person fit is caused by cheating or guessing. Table 3 shows the criteria for assessing the items and the students' ability levels.

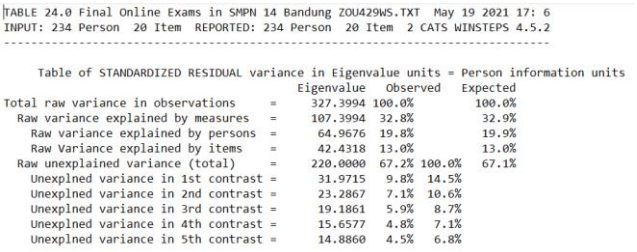
**Table 3.** The reference table for the items' validity and students' abilities [32].

| Statistics | Criteria | Additional information |
|---|---|---|
| Point measure Correlation (PTMEA-CORR) | 0.4-0.85 | To evaluate the difficulty level of the items from the hardest to the easiest |
| Model S.E | X<0.5 | X<0.5 means that it can adequately determine the students' abilities |
| Outfit Mean Square Values (MNSQ) | 0.5 <X<1.5 | A too-large MNSQ point means that students with high ability can answer incorrectly<br>A too-small MNSQ point means that students with low abilities can answer the items correctly |
| Outfit Z-standartized Values (ZSTD) | -2.0 <ZSTD <+2.0 | . |

## *Discussion*

### 1. Validity

As shown in Figure 3, the analysis of the items' validity was performed with the Rasch model, and the construct and content validity results obtained were organized in the output tables of item dimensionality.



```
TABLE 24.0 Final Online Exams in SMPN 14 Bandung ZOU429WS.TXT  May 19 2021 17: 6
INPUT: 234 Person  20 Item  REPORTED: 234 Person  20 Item  2 CATS WINSTEPS 4.5.2
-------------------------------------------------------------------------------

     Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Person information units
                                    Eigenvalue   Observed   Expected
Total raw variance in observations     =     327.3994 100.0%          100.0%
  Raw variance explained by measures   =     107.3994  32.8%           32.9%
    Raw variance explained by persons  =      64.9676  19.8%           19.9%
    Raw Variance explained by items    =      42.4318  13.0%           13.0%
  Raw unexplained variance (total)     =     220.0000  67.2% 100.0%   67.1%
    Unexplned variance in 1st contrast =      31.9715   9.8%  14.5%
    Unexplned variance in 2nd contrast =      23.2867   7.1%  10.6%
    Unexplned variance in 3rd contrast =      19.1861   5.9%   8.7%
    Unexplned variance in 4th contrast =      15.6577   4.8%   7.1%
    Unexplned variance in 5th contrast =      14.8860   4.5%   6.8%
```

**Figure 3**. The output table of the construct validity analysis

The construct validity results of the *Raw variance explained* by empirical measures produced a score of 32.8%, while the Rasch model predicted 32.9%. Since the empirical construct validation is almost the same as the value predicted by the Rasch model, the minimum unidimensionality requirement of 20% was met (> 40% good, and > 60% means special).

Meanwhile, the *Unexplained variance* values were all < 15%, which is not good [33]. The construct validation was obtained with a poor assessment, as it was not performed by the teacher. During the Covid-19 pandemic, Indonesia implemented a 75% work-from-home rule, which prevented proper communication and meetings for producing and evaluating the final examination questions. The mathematics teachers experienced difficulties asking for help from other teachers to conduct the construct validation analysis while making their test. Furthermore, this was the first time the school was determining the junior high school students' graduation based on the final examination scores. Therefore, the school had a lack of experience in making adequate items to determine the junior high school student's graduation.

[32] George Engelhard and Jue Wang, 'Developing a Concept Map for Rasch Measurement Theory', *Springer Proceedings in Mathematics and Statistics*, 322 (2020), 19–29 <https://doi.org/10.1007/978-3-030-43469-4_2>.

[33] Mokshein, Ishak, and Ahmad.

**Item Fit**

Hence, construct validation can be performed by involving several validators to obtain better results. Similar items are first tested in small groups, then the results are re-evaluated per item. The teachers can also ensure that the sentences per item are unambiguous and easy to understand. However, although the construct validation was not performed, it can be predicted by the Rasch model. This model was very effectively applied, as it can obtain reliable validity analysis results and is easy to use because it can be directly analyzed by computer applications (Mair & Hatzinger, 2007). Table 4 shows the *output tables fit,* which represents the content validity analysis results that can be seen from the level of items' suitability.

**Table 4.** Item fit analysis in the mathematics Online Final Examination

TABLE 10.1 Final Online Exams in SMPN 14 Bandung ZOU429WS.TXT May 19 2021 17: 6
INPUT: 234 Person 20 Item REPORTED: 234 Person 20 Item 2 CATS WINSTEPS 4.5.2

Person: REAL SEP.: 1.61 REL.: .72 ... Item: REAL SEP.: 4.51 REL.: .95

Item STATISTICS: MISFIT ORDER

| ENTRY NUMBER | TOTAL SCORE | TOTAL COUNT | MEASURE | MODEL S.E. | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD | PTMEASUR-AL CORR. | PTMEASUR-AL EXP. | EXACT MATCH OBS% | EXACT MATCH EXP% | Item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 111 | 234 | 1.68 | .16 | 1.30 | 4.55 | 1.49 | 3.32 | A .36 | .53 | 62.7 | 71.0 | Q15 |
| 19 | 142 | 234 | .91 | .16 | 1.18 | 2.27 | 1.34 | 2.77 | B .44 | .54 | 70.9 | 74.4 | Q19 |
| 5 | 194 | 234 | -.72 | .20 | 1.09 | .75 | 1.15 | .69 | C .43 | .48 | 83.2 | 84.8 | Q5 |
| 10 | 137 | 234 | 1.04 | .16 | 1.12 | 1.67 | 1.15 | 1.31 | D .47 | .54 | 70.5 | 73.6 | Q10 |
| 17 | 162 | 234 | .37 | .17 | .99 | -.08 | 1.15 | 1.09 | E .53 | .53 | 78.6 | 78.2 | Q17 |
| 11 | 165 | 234 | .28 | .17 | 1.06 | .71 | 1.07 | .63 | F .50 | .53 | 80.0 | 78.8 | Q11 |
| 12 | 146 | 234 | .81 | .16 | 1.00 | .00 | 1.07 | .66 | G .53 | .54 | 75.5 | 75.2 | Q12 |
| 14 | 111 | 234 | 1.68 | .16 | .99 | -.18 | 1.05 | .43 | H .53 | .53 | 71.8 | 71.0 | Q14 |
| 1 | 216 | 234 | -1.91 | .27 | 1.04 | .27 | .77 | -.40 | I .37 | .37 | 91.4 | 92.2 | Q1 |
| 6 | 179 | 234 | -.16 | .18 | 1.01 | .13 | 1.02 | .15 | J .51 | .51 | 79.1 | 81.3 | Q6 |
| 3 | 198 | 234 | -.89 | .21 | .94 | -.44 | .91 | -.27 | j .49 | .46 | 88.6 | 86.0 | Q3 |
| 8 | 173 | 234 | .04 | .18 | .94 | -.59 | .83 | -1.06 | i .56 | .52 | 80.5 | 80.3 | Q8 |
| 9 | 186 | 234 | -.41 | .19 | .93 | -.62 | .82 | -.89 | h .54 | .50 | 84.1 | 82.8 | Q9 |
| 18 | 184 | 234 | -.34 | .19 | .93 | -.66 | .89 | -.54 | g .54 | .50 | 85.0 | 82.3 | Q18 |
| 4 | 201 | 234 | -1.03 | .22 | .92 | -.56 | .81 | -.64 | f .49 | .45 | 89.5 | 87.0 | Q4 |
| 16 | 183 | 234 | -.30 | .19 | .91 | -.86 | .89 | -.53 | e .55 | .51 | 84.5 | 82.1 | Q16 |
| 20 | 183 | 234 | -.30 | .19 | .91 | -.84 | .85 | -.75 | d .55 | .51 | 85.5 | 82.1 | Q20 |
| 13 | 175 | 234 | -.03 | .18 | .89 | -1.11 | .81 | -1.16 | c .58 | .52 | 84.5 | 80.6 | Q13 |
| 7 | 170 | 234 | .13 | .18 | .81 | -2.03 | .68 | -2.32 | b .63 | .53 | 83.6 | 79.7 | Q7 |
| 2 | 197 | 234 | -.85 | .21 | .80 | -1.70 | .72 | -1.14 | a .56 | .47 | 88.2 | 85.7 | Q2 |
| MEAN | 170.6 | 234.0 | .00 | .19 | .99 | .0 | .97 | .1 | | | 80.9 | 80.4 | |
| P.SD | 27.9 | .0 | .89 | .03 | .12 | 1.4 | .20 | 1.31 | | | 7.2 | 5.3 | |

Table 4 above shows the results of the item fit analysis, the level of problem difficulty, and the standard error for each item. There were two item statistics of misfit orders, namely the infit and outfit. The outfit statistics were used more frequently due to their higher sensitivity to data with extreme scores. After comparing the 20 items, 16 were found to be fit, while four were not because they did not meet the three criteria above. Table 1 shows that the topmost item, number 15, did not fit, as it does not meet the requirements for Outfit ZSTD based on the criteria above (value 3.32). This means that item 15 should be investigated more because it does not contribute adequately to classifying students' mathematical abilities. Another interpretation is that the questions were too difficult for students, and thus cannot classify those with high and low mathematical abilities. The next step was to revise or discard the test item. Conversely, the infit mean square statistic showed that the average of each item was 1.0, meaning it was within acceptable limits. The MNSQ outfits for test items 19 and 7 were outside the acceptable limit, while number 1 had a low measure correlation, meaning it was too easy or difficult for students. Therefore, the final decision on the test items should be investigated and revised again.

**2. Reliability**

The Online mathematics final examination had 20 items divided into four subtopics, namely Numbers, Algebra, Geometry and Measurement, Probability, and Statistics. Subsequently, the Rasch model analyzed the correlation between the students' mathematical abilities and the test items, using an examination taken by 234 students. The Rasch statistical analysis model was divided into two categories, which are the 234 measured persons and the 20 measured items. Tables 5 and 6 discuss the summary statistics for each category in detail.

**Table 5.** Summary statistics of person reliability

SUMMARY OF 220 MEASURED (NON-EXTREME) Person

| | TOTAL SCORE | COUNT | MEASURE | MODEL S.E. | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
|---|---|---|---|---|---|---|---|---|
| MEAN | 14.2 | 20.0 | 1.28 | .64 | .99 | .10 | .97 | .13 |
| SEM | .3 | .0 | .09 | .01 | .01 | .04 | .03 | .05 |
| P.SD | 4.4 | .0 | 1.35 | .15 | .17 | .66 | .45 | .79 |
| S.SD | 4.4 | .0 | 1.35 | .16 | .17 | .66 | .45 | .80 |
| MAX. | 19.0 | 20.0 | 3.29 | 1.05 | 1.47 | 2.21 | 3.31 | 2.53 |
| MIN. | 2.0 | 20.0 | -2.48 | .48 | .62 | -1.88 | .30 | -1.66 |

REAL RMSE .68 TRUE SD 1.16 SEPARATION 1.70 Person RELIABILITY .74
MODEL RMSE .66 TRUE SD 1.17 SEPARATION 1.77 Person RELIABILITY .76
S.E. OF Person MEAN = .09

Person RAW SCORE-TO-MEASURE CORRELATION = .96
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .86

Table 5 is a statistical summary of the 234 students who took the Online Final Mathematics Examination. Cronbach's alpha, which is the person's raw score reliability that measures reliability using the interaction between the person and the item as a whole, was 0.86. According to the rating scale of the items' quality criteria, the Cronbach's alpha fell in the range of 0.81 to 0.90, meaning the overall quality of the items was in the excellent category. The person measure was +1.28 logit, showing the average value of all students working on the item given. Meanwhile, the average value was larger than the logit value of 0.0, indicating a tendency for students' mathematical abilities to be higher than the level of difficulty. A value of 0.74 was obtained for the person reliability, indicating that the consistency of the students' answers was in the sufficient or acceptable category. The INFIT MNSQ and OUTFIT MNSQ produced average values of 0.99 and 0.97, respectively, where the ideal value is 1, i.e., the closer to 1, the better. In addition, the INFIT ZSTD and OUTFIT ZSTD gave average table person values of 0.10 and 0.13, where, in this case, the ideal value is 0.0, that is, a value closer to 0.0 depicts better quality. The value of separation was 1.7, meaning it can separate the students' mathematical abilities into three groups, namely low, medium, and high. Here, a larger separation value signifies a better quality of the items for dividing the students' mathematical abilities.

**Table 6.** Summary statistics for item reliability

SUMMARY OF 20 MEASURED (NON-EXTREME) Item

| | TOTAL SCORE | COUNT | MODEL MEASURE | MODEL S.E. | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
|---|---|---|---|---|---|---|---|---|
| MEAN | 170.6 | 234.0 | .00 | .19 | .99 | .03 | .97 | .06 |
| SEM | 6.4 | .0 | .20 | .01 | .03 | .33 | .05 | .30 |
| P.SD | 27.9 | .0 | .89 | .03 | .12 | 1.44 | .20 | 1.32 |
| S.SD | 28.6 | .0 | .91 | .03 | .12 | 1.47 | .21 | 1.35 |
| MAX. | 216.0 | 234.0 | 1.68 | .27 | 1.30 | 4.55 | 1.49 | 3.32 |
| MIN. | 111.0 | 234.0 | -1.91 | .16 | .80 | -2.03 | .68 | -2.32 |
| REAL RMSE | .19 | TRUE SD | .86 | SEPARATION | 4.51 | Item | RELIABILITY | .95 |
| MODEL RMSE | .19 | TRUE SD | .87 | SEPARATION | 4.59 | Item | RELIABILITY | .95 |
| S.E. OF Item MEAN = .20 | | | | | | | | |

Table 6 shows summary statistics for the 20 items of the Online Final Mathematics Examination. The statistical data results were aimed at analyzing the item categories, either difficult, medium, or easy, and provide overall quality information of the student response patterns, the instrument used, and the interaction between Pearson and the items. Subsequently, the average INFIT MNSQ and OUTFIT MNSQ values were 0.99 and 0.97, where the ideal value is 1, that is, the closer to 1, the better. Conversely, the average INFIT ZSTD and OUTFIT ZSTD values of the item table were 0.03 and 0.06, and in this case, the ideal value was 0.0, where a closer score to 0.0 signifies better quality. The item separation was 4.59, indicating that the items can be divided into 5 categories, namely very easy, easy, medium, difficult, and very difficult. Based on the data analysis results, the difficulty levels were obtained from the *item measure output* results. Then, the items were grouped by combining the logit mean and the standard deviation value, which produced averages of 0.00 and 0.91, respectively. These values were used to identify the separation.

## The Wright Maps

The right section of Figure 4 shows the distribution of the problem difficulty level, while the left shows the students' mathematical abilities with the same logit ruler [34]. Subsequently, the lowest student ability level was seen from the Min logit person at -2.48, the highest was the max person logit at +3.29, and the average was +1.28. The average logit value was positive and higher than the M code (see Figure 4), meaning the average mathematical ability of the students was in the moderate and 'quite good' group. Table 7 shows the specific classification of the students' abilities, which were

[34] William J. Boone, Melissa S. Yale, and John R. Staver, *Rasch Analysis in the Human Sciences* (Springer, 2014) <https://doi.org/10.1007/978-94-007-6857-4>.

more in the moderate and very good groups than the low and very low groups.

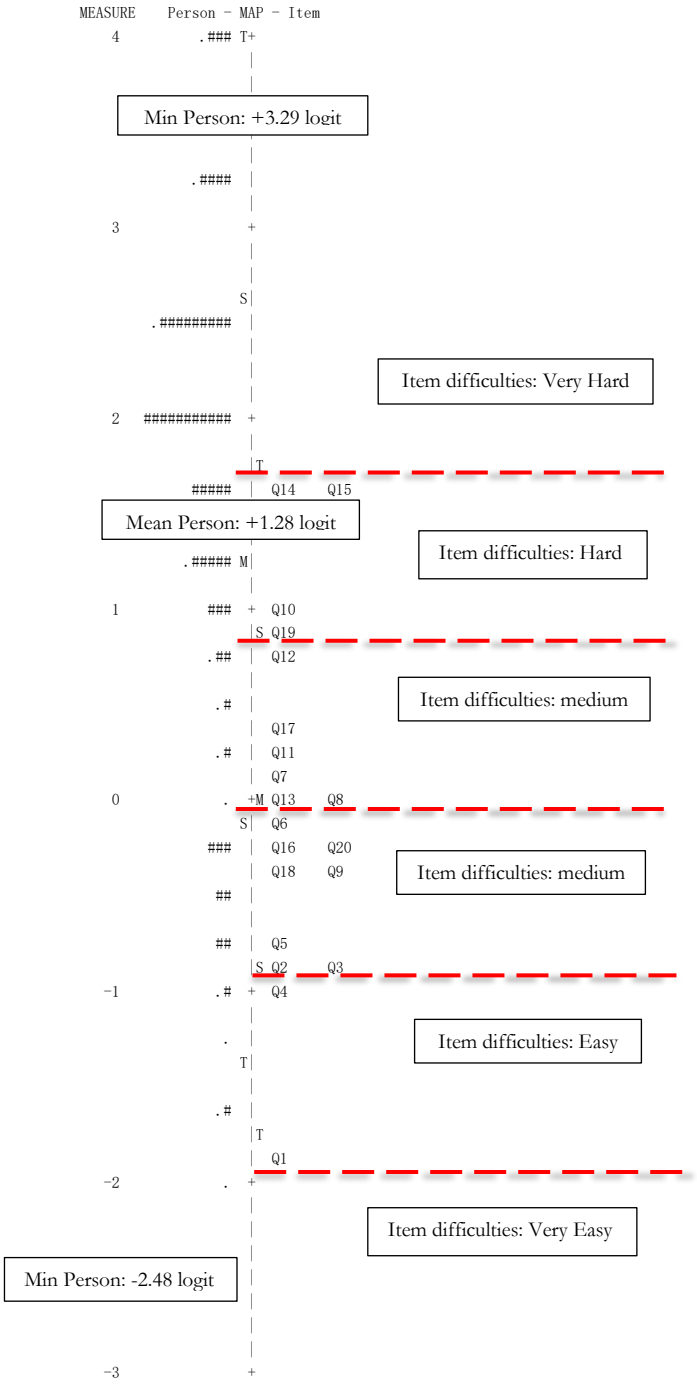



**Figure 4.** Person-item distribution map

**Table 7.** Classification of the students' mathematical abilities

| Code of classification | Code of students | Total of student | Interpretation |
|---|---|---|---|
| +T | 002IX 012XI 019IX 028IX 038IX 043IX 054IX 079IX 084IX106IX 167IX 175IX 203IX 218IX | 14 | Very high ability student |
| +S | 010IX 015IX 020IX 034IX 046IX 048IX 091IX 109IX 117IX 118IX 121IX 125IX 126IX 128IX 131IX 132IX 136IX 141IX 144IX 145IX 149IX 156IX 161IX 162IX 171IX 183IX 184IX 185IX 191IX 193IX 197IX 198IX 201IX 215IX 219IX 222IX 226IX | 37 | High ability student |
| M | 004IX 013IX 014IX 022IX 045IX 053IX 058IX 059IX 060IX 064IX 067IX 071IX 072IX 074IX 077IX 080IX 090IX 098IX 099IX 100IX 103IX 105IX 110IX 116IX 119IX 127IX 134IX 137IX 147IX 153IX 155IX 160IX 164IX 170IX 188IX 205IX 208IX 210IX 212IX 217IX 221IX 224IX 225IX 231IX 006IX 024IX 033IX 042IX 047IX 070IX 096IX 102IX 104IX 113IX 140IX 146IX 150IX 154IX 166IX 187IX 199IX 216IX 227IX 001IX 005IX 009IX 011IX 021IX 051IX 057IX 069IX 097IX 111IX 139IX 142IX 143IX 148IX 151IX 169IX 178IX 194IX202IX 206IX 234IX 016IX 027IX 035IX 062IX 073IX 087IX 130IX 135IX 163IX214IX 228IX 229IX 049IX 092IX 133IX 165IX 168IX 177IX 186IX 189IX 200IX 008IX 101IX 158IX 172IX 176IX 207IX 036IX 083IX 088IX 192IX 230IX 052IX 094IX 204IX | 138 | moderate ability student |
| -S | 003IX 007IX 025IX 066IX 076IX 085IX 086IX 107IX 181IX 182IX 232IX 233IX 023IX 039IX 040IX 082IX 129IX 138IX 159IX 180IX 030IX 037IX 041IX 055IX 081IX 089IX 179IX 211IX 018IX 031IX 152IX 173IX 209IX 115IX 123IX | 35 | Low ability Student |
| -T | 061IX 108IX 190IX 195IX 196IX 029IX 124IX 032IX 093IX 112IX | 10 | Very low ability student |

**Data of the students' answers using the Guttman Scalogram**

The Guttman scalogram observed the student answers on each item and ranked their abilities from the highest to the lowest [35]. As shown in Figure 5, five people in class 9A, 3 in 9B as well as 2 in 9C, 9D, and 9H achieved perfect scores. These 14 students answered all the items on the mathematics Online Final Examination correctly, while the students with codes 032IXG, 093IXH, and 112IXE had the lowest abilities. The three lowest students were at the bottom and could only answer two items out of the 20 correctly.

GUTTMAN SCALOGRAM OF RESPONSES:
Person |Item
| 112 1 1111111
|14325986063871729045
|--------------------
2 +11111111111111111111 002IXB
12 +11111111111111111111 012XIA
19 +11111111111111111111 019IXB
28 +11111111111111111111 028IXA
38 +11111111111111111111 038IXD
43 +11111111111111111111 043IXD
54 +11111111111111111111 054IXB
79 +11111111111111111111 079IXA
84 +11111111111111111111 084IXA
106 +11111111111111111111 106IXC
167 +11111111111111111111 167IXH
175 +11111111111111111111 175IXH
203 +11111111111111111111 203IXC
218 +11111111111111111111 218IXA
17 +11111111111111111110 017IXD
26 +11111111111111111110 026IXC
44 +11111111111111111011 044IXE
50 +11110111111111111111 050IXB
56 +11111111111111111110 056IXF
65 +11111111111111111110 065IXF
68 +11111101111111111111 068IXB
75 +11111111101111111111 075IXF
78 +11111111111111011111 078IXA
95 +11111111111111011111 095IXD
114 +11111111111111111011 114IXB
120 +11111111111111111110 120IXA
122 +11111111111111111110 122IXH
157 +11111111111111111101 157IXA
174 +11111111111111110111 174IXE
143 +11111111111001101001 143IXF
148 +11110110011111101101 148IXG
151 +11110111111100111100 151IXA
169 +11111101111101001101 169IXH
16 +11111110011011010101 016IXA
27 +11111111110110000110 027IXB
130 +11111100110110011110 130IXA
135 +10111010111101101101 135IXA
163 +11111001111111001100 163IXB
214 +11011111100111011001 214IXG
228 +11111100101111110100 228IXH
229 +10111101111110001011 229IXG
49 +11111101110011010100 049IXA
92 +11010101101111110010 092IXG
133 +11011111000111011001 133IXE
165 +11101101101001100111 165IXC
168 +10111001011111110010 168IXA
177 +11110101111000110101 177IXA
186 +11111111101001100001 186IXF
189 +11010111101101001011 189IXH
200 +11111111111001010000 200IXB
8 +11111000111101000101 008IXA
101 +11101010100111100101 101IXF
158 +01110111101110000110 158IXC
172 +10011111110101011000 172IXC
176 +11000110011111110010 176IXF
207 +11111111001001001001 207IXB
36 +01111010110001110100 036IXE
83 +01011101101111000001 083IXA
88 +11111111001010001001 088IXC
87 +11111111001010001001 087IXC
230 +11101111000010001110 230IXC
52 +11111001101100000001 052IXG
94 +10111111110100000000 094IXE
204 +11011011111100000000 204IXF
3 +10101110101101000000 003IXA
7 +10110010110000101001 007IXF
25 +11011001101010001000 025IXB
66 +11011011100010000010 066IXE
76 +01111111110000000000 076IXE
85 +11110010010000000111 085IXD
86 +11111100010110000000 086IXG
107 +10110010011001011000 107IXE
181 +11111110000110000000 181IXG
182 +11111110000110000000 182IXG
232 +11100001110010101000 232IXE
233 +11110111100000100000 233IXH
23 +11101101000000001010 023IXG
39 +01100000111010001100 039IXA
40 +11101000010000011001 040IXA
82 +11100100011100100000 082IXC
129 +11001011110000000100 129IXE
138 +10100001010101100001 138IXD
159 +01101110001101000000 159IXD
180 +11111110000000100000 180IXG
30 +00110000001100001101 030IXC
37 +10001101010000100001 037IX
41 +01000010110000110001 041IXD
55 +10001001100101000100 055IXD
81 +11101001000001000100 081IXH
89 +11001110000010001000 089IXF
179 +10110010000100000101 179IXE
211 +01111000010001100000 211IXE
18 +11101000000000000101 018IXF
31 +11001010010000010000 031IXG
152 +10010000011000000101 152IXG
173 +00001110000011010000 173IXG

[35] Mokshein, Ishak, and Ahmad.

```
209 +11001000100000001010  209IXE
115 +10100000110001000000  115IXE
123 +10000101001100000000  123IXD
 61 +00110000000001001000  061IXC
108 +10000000001110000000  108IXA
190 +00010000010000101000  190IXB
195 +00000100000001010010  195IXA
196 +10000101000000010000  196IXG
 29 +01000000000000110000  029IXD
124 +10000000100000100000  124IXD
 32 +10001000000000000000  032IXG
 93 +00001000000000001000  093IXH
```

**Figure 5.** The Guttman scalogram of the students' answers

Besides sorting students' abilities from the highest to the lowest, the Guttman scalogram also sorts them from the easiest to the hardest. The item on the left was the easiest, while that on the right was the most difficult. Further analysis shows the students who were not careful when working on the examination. In the green table, students with initials 075IXF, 078IXA, and 095IXD only answered one item incorrectly out of 20 given. They worked on the last five items that had higher difficulty, indicating that their ability was sufficient, though they were not careful enough to answer the easier items. Meanwhile, the students with initials 056IXF and 065IXF in the blue box answered the first 19 items correctly but incorrectly answered number 20, showing their ability was still insufficient for the question. The Guttman scalogram can also detect students that cheat when working on the mathematics Online Final Examination. In the black box case, students with initials 088IXC and 087IXC had consecutive numbers, similar math test scores, and patterns in answering the examination questions. From the data, the teacher can further analyze and determine whether these two students cheated.

The Guttman scalogram data is very beneficial for teachers and schools. Teachers and supervisors should ensure that the Online Final Mathematics Examination is conducted honestly and that a final score without cheating is obtained. They can also remind students that attitude is more important than grades in the real world.

In addition, the Guttman scalogram data can be used by teachers to classify students' abilities. By applying this data on the Rasch model to analyze the daily or midterm test, the teacher can see the items that are still difficult and need to be explained again. The teachers can also assess students with low mathematical abilities to assist them to focus more with other learning methods or approaches. Meanwhile, teachers can provide more challenging items for the high-ability students to improve their other math skills or prepare them for competitions.

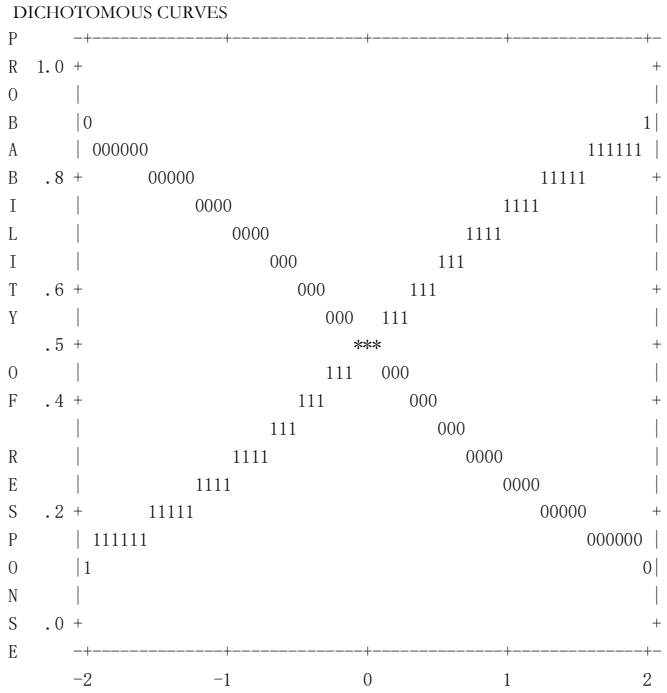



**Figure 6.** Probability of response - dichotomous curves

Figure 6 illustrates the probability categories and shows that the logit peak was around 0.9, and numbers 1 and 0 did not cover each other. The correctly and incorrectly answered items were balanced, meaning each test adequately measured and divided the students' mathematical ability and into the low and high categories. Consequently, the items can be concluded to have appropriately divided the students' abilities.

**Table 8.** Distractor analysis of the online final mathematics examination

| ENTRY NUMBER | DATA CODE | SCORE VALUE | DATA COUNT | DATA % | ABILITY MEAN | S.E. P.SD | S.E. MEAN | INFT MNSQ | OUTF MNSQ | PTMA CORR. | Item |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 14 | D | 0 | 15 | 6 | .09 | 1.34 | .36 | .7 | .6 | -.24 | Q14 |
| | B | 0 | 45 | 19 | .71 | 1.44 | .22 | 1.1 | 1.0 | -.24 | |
| | | 0 | 2 | 1 | .72 | .26 | .26 | .4 | .5 | -.05 | |
| | C | 0 | 61 | 26 | .84 | 1.18 | .15 | .9 | 1.0 | -.25 | |
| | A | 1 | 111 | 47 | 2.33 | 1.25 | .12 | 1.0 | 1.1 | .53 | |
| 15 | | 0 | 1 | 0 | -1.27 | .00 | | .1 | .1 | -.12 | Q15 |
| | A | 0 | 31 | 13 | .80 | 1.57 | .29 | 1.2 | 1.4 | -.17 | |
| | B | 0 | 46 | 20 | .98 | 1.41 | .21 | 1.2 | 1.2 | -.16 | |
| | C | 0 | 45 | 19 | 1.09 | 1.28 | .19 | 1.2 | 1.3 | -.12 | |
| | D | 1 | 111 | 47 | 2.04 | 1.42 | .14 | 1.4 | 1.6 | .36 | |
| 10 | C | 0 | 28 | 12 | .20 | 1.35 | .26 | .8 | .7 | -.31 | Q10 |
| | B | 0 | 37 | 16 | .46 | 1.43 | .24 | 1.0 | 1.1 | -.29 | |
| | | 0 | 3 | 1 | .67 | 1.61 | 1.14 | 1.4 | 1.2 | -.06 | |
| | A | 0 | 29 | 12 | 1.21 | 1.26 | .24 | 1.5 | 1.7 | -.07 | |
| | D | 1 | 137 | 59 | 2.08 | 1.28 | .11 | 1.1 | 1.1 | .47 | |
| 19 | | 0 | 1 | 0 | -.24 | .00 | | .2 | .2 | -.07 | Q19 |
| | C | 0 | 28 | 12 | .45 | 1.16 | .22 | .8 | .8 | -.25 | |
| | D | 0 | 36 | 15 | .71 | 1.22 | .21 | 1.0 | 1.2 | -.21 | |
| | A | 0 | 27 | 12 | .80 | 1.67 | .33 | 1.7 | 1.5 | -.16 | |
| | B | 1 | 142 | 61 | 2.01 | 1.38 | .12 | 1.2 | 1.6 | .44 | |
| 12 | | 0 | 1 | 0 | -.48 | .00 | | .2 | .2 | -.08 | Q12 |
| | A | 0 | 24 | 10 | -.30 | 1.14 | .24 | .5 | .4 | -.39 | |
| | D | 0 | 20 | 9 | .29 | 1.08 | .25 | .7 | .7 | -.24 | |
| | B | 0 | 43 | 18 | .91 | 1.28 | .20 | 1.3 | 1.4 | -.17 | |
| | C | 1 | 146 | 62 | 2.10 | 1.28 | .11 | 1.0 | 1.2 | .53 | |
| 17 | D | 0 | 14 | 6 | -.03 | 1.22 | .34 | .7 | .6 | -.25 | Q17 |
| | A | 0 | 29 | 12 | .32 | 1.37 | .26 | 1.0 | 1.3 | -.28 | |
| | C | 0 | 29 | 12 | .37 | 1.13 | .21 | .9 | 1.0 | -.27 | |
| | B | 1 | 162 | 69 | 2.01 | 1.31 | .10 | 1.1 | 1.4 | .53 | |
| 11 | | 0 | 2 | 1 | -.92 | .68 | .68 | .3 | .2 | -.15 | Q11 |
| | C | 0 | 24 | 10 | -.16 | 1.19 | .25 | .7 | .6 | -.36 | |
| | D | 0 | 14 | 6 | -.06 | 1.11 | .31 | .7 | .6 | -.25 | |
| | B | 0 | 29 | 12 | .95 | 1.43 | .27 | 1.8 | 1.8 | -.13 | |
| | A | 1 | 165 | 71 | 1.96 | 1.29 | .10 | 1.1 | 1.1 | .50 | |
| 7 | C | 0 | 24 | 10 | -.13 | 1.42 | .30 | .9 | .8 | -.36 | Q7 |
| | B | 0 | 25 | 11 | -.12 | 1.14 | .23 | .7 | .6 | -.36 | |
| | A | 0 | 15 | 6 | .04 | .73 | .20 | .6 | .5 | -.25 | |
| | D | 1 | 170 | 73 | 2.06 | 1.18 | .09 | .8 | .8 | .63 | |
| 8 | D | 0 | 7 | 3 | -.56 | 1.26 | .51 | .6 | .4 | -.24 | Q8 |
| | | 0 | 3 | 1 | -.50 | 1.12 | .79 | .6 | .4 | -.15 | |
| | B | 0 | 22 | 9 | -.26 | 1.14 | .25 | .7 | .5 | -.37 | |
| | A | 0 | 29 | 12 | .46 | 1.27 | .24 | 1.2 | 1.2 | -.25 | |
| | C | 1 | 173 | 74 | 1.98 | 1.25 | .10 | 1.0 | .9 | .56 | |
| 13 | D | 0 | 13 | 6 | -.66 | 1.13 | .33 | .5 | .4 | -.34 | Q13 |
| | C | 0 | 18 | 8 | -.02 | 1.23 | .30 | .9 | .7 | -.28 | |
| | B | 0 | 28 | 12 | .24 | 1.29 | .25 | 1.0 | 1.0 | -.30 | |
| | A | 1 | 175 | 75 | 1.98 | 1.23 | .09 | .9 | .9 | .58 | |
| 6 | | 0 | 1 | 0 | -2.48 | .00 | | .0 | .0 | -.17 | Q6 |
| | A | 0 | 19 | 8 | -.48 | 1.21 | .29 | .7 | .5 | -.38 | |
| | D | 0 | 8 | 3 | .22 | 1.53 | .58 | 1.0 | 1.9 | -.16 | |
| | C | 0 | 27 | 12 | .50 | 1.17 | .23 | 1.2 | 1.2 | -.23 | |
| | B | 1 | 179 | 76 | 1.90 | 1.29 | .10 | 1.0 | 1.0 | .51 | |
| 16 | C | 0 | 18 | 8 | -.29 | 1.44 | .35 | .9 | .8 | -.33 | Q16 |
| | D | 0 | 20 | 9 | -.02 | 1.14 | .26 | .9 | .8 | -.30 | |
| | B | 0 | 13 | 6 | .00 | 1.49 | .43 | 1.0 | 1.2 | -.23 | |
| | A | 1 | 183 | 78 | 1.91 | 1.25 | .09 | .9 | .9 | .55 | |
| 20 | D | 0 | 13 | 6 | -.37 | 1.06 | .31 | .7 | .6 | -.29 | Q20 |
| | C | 0 | 16 | 7 | -.23 | 1.21 | .31 | .8 | .6 | -.30 | |
| | B | 0 | 22 | 9 | .12 | 1.46 | .32 | 1.1 | 1.2 | -.29 | |
| | A | 1 | 183 | 78 | 1.91 | 1.26 | .09 | .9 | .9 | .55 | |
| 18 | D | 0 | 9 | 4 | -.53 | 1.24 | .44 | .7 | .5 | -.26 | Q18 |
| | B | 0 | 16 | 7 | -.26 | 1.16 | .30 | .8 | .6 | -.31 | |
| | C | 0 | 24 | 10 | .12 | 1.46 | .30 | 1.2 | 1.2 | -.30 | |
| | | 0 | 1 | 0 | .23 | .00 | | .8 | .5 | -.05 | |
| | A | 1 | 184 | 79 | 1.90 | 1.26 | .09 | .9 | .8 | .54 | |
| 9 | A | 0 | 16 | 7 | -.49 | 1.20 | .31 | .7 | .5 | -.35 | Q9 |
| | C | 0 | 17 | 7 | -.10 | 1.13 | .28 | .8 | .7 | -.29 | |
| | B | 0 | 12 | 5 | .15 | 1.45 | .44 | 1.3 | 1.2 | -.20 | |
| | | 0 | 3 | 1 | .25 | 1.27 | .90 | 1.1 | 1.2 | -.09 | |
| | D | 1 | 186 | 79 | 1.89 | 1.28 | .09 | .9 | .9 | .54 | |
| 5 | C | 0 | 6 | 3 | -.59 | 1.13 | .50 | .7 | .5 | -.22 | Q5 |
| | B | 0 | 13 | 6 | -.54 | 1.22 | .35 | .8 | .5 | -.32 | |
| | D | 0 | 18 | 8 | .45 | 1.21 | .29 | 1.5 | 1.3 | -.19 | |
| | | 0 | 3 | 1 | 1.09 | 1.57 | 1.11 | 1.7 | 4.4 | -.03 | |
| | A | 1 | 194 | 83 | 1.77 | 1.37 | .10 | 1.1 | 1.2 | .43 | |
| 2 | B | 0 | 10 | 4 | -.93 | 1.29 | .43 | .6 | .5 | -.33 | Q2 |
| | A | 0 | 19 | 8 | -.62 | .86 | .20 | .6 | .4 | -.41 | |
| | D | 0 | 8 | 3 | .27 | 1.52 | .58 | 1.2 | 1.7 | -.15 | |
| | C | 1 | 197 | 84 | 1.85 | 1.26 | .09 | .8 | .8 | .56 | |
| 3 | D | 0 | 5 | 2 | -1.11 | 1.48 | .74 | .6 | .6 | -.25 | Q3 |
| | C | 0 | 15 | 6 | -.43 | .97 | .26 | .8 | .5 | -.33 | |
| | B | 0 | 14 | 6 | -.01 | 1.52 | .42 | 1.2 | 1.2 | -.25 | |
| | | 0 | 2 | 1 | 1.24 | 1.25 | 1.25 | 2.4 | 3.1 | -.01 | |
| | A | 1 | 198 | 85 | 1.79 | 1.32 | .09 | .9 | .9 | .49 | |
| 4 | B | 0 | 11 | 5 | -.69 | 1.13 | .36 | .7 | .5 | -.32 | Q4 |
| | A | 0 | 7 | 3 | -.42 | 1.37 | .56 | .9 | .8 | -.22 | |
| | C | 0 | 15 | 6 | -.09 | 1.41 | .38 | 1.2 | 1.0 | -.27 | |
| | D | 1 | 201 | 86 | 1.77 | 1.33 | .09 | .9 | .9 | .49 | |
| 1 | C | 0 | 10 | 4 | -.99 | 1.03 | .34 | .7 | .4 | -.34 | Q1 |
| | D | 0 | 2 | 1 | -.25 | .48 | .48 | 1.3 | .5 | -.11 | |
| | A | 0 | 6 | 3 | .32 | 1.01 | .45 | 1.9 | 1.4 | -.12 | |
| | B | 1 | 216 | 92 | 1.63 | 1.43 | .10 | 1.1 | 1.1 | .37 | |

Table 8 shows the distractor analysis of the 20 items of the online final mathematics examination evaluated by the Rasch model. From the data distractor analysis table, the count data and the number of students that answered the multiple choices incorrectly and correctly can be seen. Meanwhile, the up or down tendency of the ability mean shows whether the distractor analysis is working properly [36]. When this score is down, the distractor analysis is not going well and the teacher can revise the answer choices for the items and retest them. Based on the data in Table 8, the 20 mathematics Online Final Examination items have good distractor analyses.

[36] Stefanie A. Wind and Jessica D. Gale, 'Diagnostic Opportunities Using Rasch Measurement in the Context of a Misconceptions-Based Physical Science Assessment', *Science Education*, 99.4 (2015), 721–41 <https://doi.org/10.1002/sce.21172>.

The final examinations in grades 9 and 12 are very important for determining the students' ability level in schools, cities, and provinces [37]. According to the results, the government can plan to improve the schools' quality in each province [38]. The schools can measure their students' abilities quality and the teachers can make evaluations on lesson plans and teaching methods, with the goal of improving the quality of education in Indonesia [39]. However, the government has not held a national examination for two years, particularly in 2020 and 2021, and has entrusted each school to make its final examination to determine students' graduation. Therefore, final examination items should be made appropriately and professionally and each item should be carefully evaluated to ensure the students' mathematical abilities are measured properly. In addition, the items should be fairly easy to allow the students to pass with good grades and slightly difficult to test their abilities. The items are fixed according to the reference, comprising of 30% easy questions, 50% moderate, and 20% at a difficult level.

Our study found that most students did well on the test. this is indicated by the positive mean logit (+1.28). A logit value close to 0 indicates that the final examination test is good and can measure student mathematics achievement. Table 7 shows that students with high ability are more than students with low ability. This can be explained by the fact that the sample in this study is a public school with the accreditation level of "A." This may explain why the children in the school have good mathematics achievement.

Subsequently, the research result shows the importance of validating the final examination items. This indicates that although public schools had the level "A" accreditation, the construct validity results did not reach the minimum score. Furthermore, the Rasch model found several items that were not good, and four out of the 20 items needed to be evaluated and revised because they were outside the fit item criteria. Hence, these four items require further revision and analysis to correspond with the fit category. In this study, 14 questions had a medium level of difficulty, four were difficult, and two were easy. However, seeing that the ability of students is quite good when working on questions, perhaps the school can increase the level of difficulty of the test items.

Cheating is an act that should not be done by students, especially in the final examinations. Online-based examinations create new challenges for schools and teachers to ensure that each student work on questions honestly and independently. In this study, the Guttman scalogram is used to find out in detail each student's answer. It can identify indicators of student cheating when working on practice questions. Guttman scalogram in Rasch

[37] Y Geng and others, 'Parental Care and Depressive Symptoms among Chinese Medical Students: Roles of Empathy and Gender', *BMC Medical Education*, 22.1 (2022) <https://doi.org/10.1186/s12909-022-03524-2>; Z Lin and others, 'Short-Term Effects of Personal Exposure to Temperature Variability on Cardiorespiratory Health Based on Subclinical Non-Invasive Biomarkers', *Science of the Total Environment*, 843 (2022) <https://doi.org/10.1016/j.scitotenv.2022.157000>; D Spence and others, 'Use of Resources and Method of Proctoring During the NBCRNA Continued Professional Certification Assessment: Analysis of Outcomes', *Journal of Nursing Regulation*, 10.3 (2019), 37–46 <https://doi.org/10.1016/S2155-8256(19)30147-4>.

[38] Candace Walkington, 'Design Research on Personalized Probem Posing in Algebra', *Proceedings of the 39th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, 2017, 195–202; Sufen Xu and Yun Wang, 'Problems and Suggestions on Reform of Teacher Qualification Examination under the Background of National Unified Examination', in *Advances in Social Science, Education and Humanities Research*, 2018, CCXLVI, 314–17 <https://doi.org/10.2991/icpel-18.2018.74>; Wu, Zhou, and Mo.

[39] Ujian Nasional, U N Harapan, and Gunadi H Sulistyo, 'Ujian Nasional ( UN ): Harapan , Tantangan , Dan Peluang', *Wacana*, 9.1 (2015) <https://doi.org/10.17510/wjhi.v9i1.224>; Chatarina Sitoresmi Triwiniastuti and others, 'Implementasi Program Ujian Nasional Di SMA Negeri', *Kelola: Jurnal Manajemen Pendidikan*, 20, 2019, 11–12.

model has been widely used to find out each student's ability in detail [40].

This research contributes to the importance of preparing the final examination items appropriately and seriously. It also explains the steps of using the Rasch model to analyze important information needed for preparing the final examination. First, it determines whether the construct validity is in a good category (X>40%). Second, it investigates the misfits on the statistical item table to observe the difficulty level of the items as a whole and analyze the personal statistics to evaluate the students' overall abilities. Third, it uses the Wright Maps table to compare the abilities of each student and the difficulty level of each item in more detail. Fourth, it analyzes the Guttman scalogram to observe the students' ability to answer the items in detail. In conclusion, these steps ensure that all items are well-prepared and whether the test is unidimensional. It also ascertains that there is no sentence error evidence, luck in answering, distractor analysis error, miscoding, etc. This research concludes that the Rasch measurement model can be used by students, teachers, and schools to adequately prepare the final examination items.

## Conclusion

Indonesia stopped holding national examinations, forcing each school to prepare its own final examination items to determine the students' graduation. By processing the final examination data using the Rasch measurement model, the items and the students' work can be analyzed. The model can show data and help teachers and schools analyze the students' abilities in working on and answering each test item on the online final mathematics examination. Furthermore, the Rasch model shows the difficulty level of each item to improve the its quality. Therefore, this model can enhance the achievement of item results and determine better goals, meaning the online final examination distractor analysis in math courses works well.

Also, the findings show that the Rasch model can analyze the mathematics quality at public schools in Indonesia. The results of the overall construct validation show that the items were not well prepared, and after further investigation, four items out of the 20 were discovered not to meet the fitness criteria. This means the four items need to be further observed, revised, or discarded. Also, the results of the person and item reliability show that the items were in the 'excellent' category. The students' abilities to work on the online final examination was above average. In addition, the *Guttman scalogram of responses* can show data on the student abilities from the highest to the lowest, as well as the less careful students, the ability limits, and those who likely cheated.

The research shows that the teachers should prepare well to achieve good construct validity when examining students, especially on important examinations, such as UTS and UAS. It also recommended that teachers use the Rasch model to obtain more specific items analysis. The model can help produce more objective examination items that can work well to measure the students' abilities. Another important message is that teachers should focus on students with low mathematical abilities and help them improve their learning outcomes.

[40] P. Susongko, 'Validation of Science Achievement Test with the Rasch Model', *Jurnal Pendidikan IPA Indonesia*, 5.2 (2016), 268–77 <https://doi.org/10.15294/jpii.v5i2.7690>; Der Ching Yang and Iwan Andi Sianturi, 'An Analysis of Singaporean versus Indonesian Textbooks Based on Trigonometry Content', *Eurasia Journal of Mathematics, Science and Technology Education*, 13.7 (2017), 3829–48 <https://doi.org/10.12973/eurasia.2017.00760a>; Valeriu Frunzaru and Nicoleta Corbu, 'Students' Attitudes towards Knowledge and the Future of Work', *Kybernetes*, 49.7 (2020), 1987–2002 <https://doi.org/10.1108/K-07-2019-0512>.

**Limitations and suggestions for further research**

There were several limitations to this research. First, it was only conducted with a small scale of 234 students, and only one public school with a good accreditation ranking in the city of Bandung was selected. Hence, the preparation of the final examination items for the 9th-grade junior high school yielded quite good results. However, analysis of public schools in remote regions may get different results, and further research should also analyze the final examinations in private schools. The second limitation is that only the final examination items at the junior high were analyzed, hence other school levels need to be investigated. Third, only math examination items were analyzed, while the final examination to determine student graduation consists of many subjects such as science, foreign language, and others. Therefore, further research can analyze other subjects tested on the final examination using the Rasch model.

**Declaration of interest**

No conflict of interest was declared. The authors are responsible for the content and writing of this article.

**References**

***Books***

***Journals***


Aini, Rahmi Qurota, Arif Rachmatullah, and Minsu Ha, 'Indonesian Primary School and Middle School Students' Attitudes toward Science: Focus on Gender and Academic Level', *Journal of Baltic Science Education*, 18.5 (2019), 654–67 <https://doi.org/10.33225/jbse/19.18.654>

Bagiyono, 'Analisis Tingkat Kesukaran Dan Daya Pembeda Sial Ujian Pelatihan Radiografi Tingkat 1', *Widyanuklida*, 16.No. 1 (2017), 1–12

Bai, Chong-en, Wei Chi, and Xiaoye Qian, 'China Economic Review Do College Entrance Examination Scores Predict Undergraduate GPAs ? A Tale of Two Universities', *China Economic Review*, 30 (2014), 632–47 <https://doi.org/10.1016/j.chieco.2013.08.005>

Boone, William J., Melissa S. Yale, and John R. Staver, *Rasch Analysis in the Human Sciences* (Springer, 2014) <https://doi.org/10.1007/978-94-007-6857-4>

Chai, Ching Sing, Xingwei Wang, and Chang Xu, 'An Extended Theory of Planned Behavior for the Modelling of Chinese Secondary School Students' Intention to Learn Artificial Intelligence', *Mathematics*, 8.11 (2020), 1–18 <https://doi.org/10.3390/math8112089>

Engelhard, George, and Jue Wang, 'Developing a Concept Map for Rasch Measurement Theory', *Springer Proceedings in Mathematics and Statistics*, 322 (2020), 19–29 <https://doi.org/10.1007/978-3-030-43469-4_2>

Faradillah, Ayu, and Leha Febriani, 'Mathematical Trauma Students' Junior High School Based on Grade and Gender', *Infinity Journal*, 10.1 (2021), 53 <https://doi.org/10.22460/infinity.v10i1.p53-68>

Frunzaru, Valeriu, and Nicoleta Corbu, 'Students' Attitudes towards Knowledge and the Future of Work', *Kybernetes*, 49.7 (2020), 1987–2002 <https://doi.org/10.1108/K-07-2019-0512>

Gan, Lijuan, and Tommy Tanu Wijaya, 'MEASURING STUDENT'S LOGICAL REASONING SKILLS OF CHINESE SENIOR HIGH SCHOOL USING RASCH MEASUREMENT MODEL', *JURNAL MathEdu (Mathematic Education Journal)*, 4.3 (2021), 142–49

Geng, Y, W Fei, Z Tang, S Wang, J Yu, M Zhang, and others, 'Parental Care and Depressive Symptoms among Chinese Medical Students: Roles of Empathy and Gender', *BMC Medical Education*, 22.1 (2022) <https://doi.org/10.1186/s12909-022-03524-2>

Hardouin, Jean Benoit, 'Rasch Analysis: Estimation and Tests with Raschtest', *Stata Journal*, 7.1 (2007), 22–44 <https://doi.org/10.1177/1536867x0700700102>

Haspel, Richard L., Yulia Lin, Patrick Fisher, Asma Ali, and Eric Parks, 'Development of a

Validated Exam to Assess Physician Transfusion Medicine Knowledge', *Transfusion*, 54.5 (2014), 1225–30 <https://doi.org/10.1111/trf.12425>

Hermita, N., A. Suhandi, E. Syaodih, A. Samsudin, H. Marhadi, S. Sapriadil, and others, 'Level Conceptual Change Pre-Service Elementary Teachers on Electric Current Conceptions through Visual Multimedia Supported Conceptual Change', *Journal of Physics: Conference Series*, 1013.1 (2018), 1–8 <https://doi.org/10.1088/1742-6596/1013/1/012060>

Hohne, H. H., 'The Prediction of Academic Success', *Australian Journal of Psychology*, 1.1 (1949), 38–42 <https://doi.org/10.1080/00049534908256014>

Huang, Liu, Fan Huang, Pey Tee Oon, and Miranda Chi Kuan Mak, 'Constructs Evaluation of Student Attitudes towards Science', *Eurasia Journal of Mathematics, Science and Technology Education*, 15.12 (2019) <https://doi.org/10.29333/ejmste/109168>

Jennings, Nicholas B., Marion K. Slack, Lea E. Mollon, and Terri L. Warholak, 'Measurement Characteristics of a Concept Classification Exam Using Multiple Case Examples: A Rasch Analysis', *Currents in Pharmacy Teaching and Learning*, 8.1 (2016), 31–38 <https://doi.org/10.1016/j.cptl.2015.09.010>

Karlin, Omar, and Sayaka Karlin, 'Making Better Tests with the Rasch Measurement Model', *InSight: A Journal of Scholarly Teaching*, 13.1 (2018), 76–100 <https://doi.org/10.46504/14201805ka>

Li, Mengzhou, Lei Luo, Sujoy Sikdar, Navid Ibtehaj Nizam, Shan Gao, Hongming Shan, and others, 'Optimized Collusion Prevention for Online Exams during Social Distancing', *Npj Science of Learning*, 6.1 (2021) <https://doi.org/10.1038/s41539-020-00083-3>

Lin, Z, L Yang, P Chen, T Wei, J Zhang, Y Wang, and others, 'Short-Term Effects of Personal Exposure to Temperature Variability on Cardiorespiratory Health Based on Subclinical Non-Invasive Biomarkers', *Science of the Total Environment*, 843 (2022) <https://doi.org/10.1016/j.scitotenv.2022.157000>

Locastro, Virginia, 'The English in Japanese University Entrance Examinations: A Sociocultural Analysis', *World Englishes*, 9.3 (1990), 343–54 <https://doi.org/10.1111/j.1467-971X.1990.tb00271.x>

Manullang, Sardjana Orba, Erwinsyah Satria, Universitas Krisnadwipayana, and Universitas Bung Hatta, 'The Review of the International Voices on the Responses of the Worldwide School Closures Policy Searching during Covid-19 Pandemic', 5.2 (2020), 1–13

Mokshein, Siti Eshah, Haliza Ishak, and Hishamuddin Ahmad, 'The Use of Rasch Measurement Model in English Testing', *Cakrawala Pendidikan*, 38.1 (2019), 16–32 <https://doi.org/10.21831/cp.v38i1.22750>

Nasional, Ujian, U N Harapan, and Gunadi H Sulistyo, 'Ujian Nasional ( UN ): Harapan , Tantangan , Dan Peluang', *Wacana*, 9.1 (2015) <https://doi.org/10.17510/wjhi.v9i1.224>

Nopiah, Zulkifli Mohd, Nur Arzilah Ismail, Haliza Othman, Izamarlina Asshaari, Noorhelyna Razali, Mohd Haniff Osman, and others, 'Identification of Student Achievement and Academic Profile in the Linear Algebra Course: An Analysis Using the Rasch Model', *2011 3rd International Congress on Engineering Education: Rethinking Engineering Education, The Way Forward, ICEED 2011*, 2011, 197–202 <https://doi.org/10.1109/ICEED.2011.6235389>

Othman, Haliza, Izamarlina Asshaari, Hafizah Bahaludin, Zulkifli Mohd Nopiah, and Nur Arzilah Ismail, 'Application of Rasch Measurement Model in Reliability and Quality Evaluation of Examination Paper for Engineering Mathematics Courses', *Procedia - Social and Behavioral Sciences*, 60.2009 (2012), 163–71 <https://doi.org/10.1016/j.sbspro.2012.09.363>

Pandey, Digvijay, Gabriel A. Ogunmola, Wegayehu Enbeyle, Marzuk Abdullahi, Binay Kumar Pandey, and Sabyasachi Pramanik, 'COVID-19: A Framework for Effective Delivering of Online Classes During Lockdown', *Human*

*Arenas*, 0123456789, 2021 <https://doi.org/10.1007/s42087-020-00175-x>

Prasetyanto, Dwi, Muhamad Rizki, and Yos Sunitiyoso, 'Online Learning Participation Intention after COVID-19 Pandemic in Indonesia: Do Students Still Make Trips for Online Class?', *Sustainability (Switzerland)*, 14.4 (2022) <https://doi.org/10.3390/su14041982>

Purwanto, Agus, Rudy Pramono, M Asbari, P B Santoso, L M Wijayanti, C C Hyun, and others, 'Studi Eksploratif Dampak Pandemi COVID-19 Terhadap Proses Pembelajaran Online Di Sekolah Dasar', *EduPsyCouns Journal*, 2.1 (2020)

Putra, Purniadi, Fahrina Yustiasari Liriwati, and Tasdin Tahrim, 'The Students Learning from Home Experience during Covid-19 School Closures Policy in Indonesia', *Jurnal Iqra' : Kajian Ilmu Pendidikan*, 5.2 (2020), 30–42

Putra, Zetra Hainul, Gustimal Witri, and Intan Kartika Sari, 'Prospective Elementary Teachers' Perspectives on Online Mathematics Learning during Coronavirus Outbreak', *Journal of Physics: Conference Series*, 1655.1 (2020) <https://doi.org/10.1088/1742-6596/1655/1/012057>

Retnawati, Heri, and Nidya F. Wulandari, 'The Development of Students' Mathematical Literacy Proficiency', *Problems of Education in the 21st Century*, 77.4 (2019), 502–14 <https://doi.org/10.33225/pec/19.77.502>

Sanz, Ana Gimeno, and Ana Sevilla Pavón, 'Toward Implementing Computer-Assisted Foreign Language Assessment in the Official Spanish University Entrance Examination', in *Proceedings of the 2015 EUROCALL Conference, Padova, Italy*, 2015, pp. 215–20

Shamsuddin, Hasni, and Ahmad Zamri Khairani, *Proceedings of the Regional Conference on Science, Technology and Social Sciences (RCSTSS 2016)*, *Proceedings of the Regional Conference on Science, Technology and Social Sciences (RCSTSS 2016)* (Springer Singapore, 2019) <https://doi.org/10.1007/978-981-13-0203-9>

Spence, D, R Ward, S Wooden, M Browne, H Song, R Hawkins, and others, 'Use of Resources and Method of Proctoring During the NBCRNA Continued Professional Certification Assessment: Analysis of Outcomes', *Journal of Nursing Regulation*, 10.3 (2019), 37–46 <https://doi.org/10.1016/S2155-8256(19)30147-4>

Susanto, H., A Rinaldi, and Novalia, 'Analisis Validitas Reabilitas Tingkat Kesukaran Dan Daya Beda Pada Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika', *The Journal of the Japan Society for Respiratory Endoscopy*, 6.2 (2015) <https://doi.org/10.18907/jjsre.37.3_343_4>

Susongko, P., 'Validation of Science Achievement Test with the Rasch Model', *Jurnal Pendidikan IPA Indonesia*, 5.2 (2016), 268–77 <https://doi.org/10.15294/jpii.v5i2.7690>

Suwarta, thomas harming, 'Ini Sejarah Ujian Nasional Di Indonesia', *Media Indonesia* (jakarta, 2019)

Tavakol, Mohsen, and Reg Dennick, 'Psychometric Evaluation of a Knowledge Based Examination Using Rasch Analysis: An Illustrative Guide: AMEE Guide No. 72', *Medical Teacher*, 35.1 (2013), e838–48 <https://doi.org/10.3109/0142159X.2012.737488>

Triono, Dedy, Riyanarto Sarno, and Kelly R. Sungkono, 'Item Analysis for Examination Test in the Postgraduate Student's Selection with Classical Test Theory and Rasch Measurement Model', *Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, ISemantic 2020*, 2020, 523–29 <https://doi.org/10.1109/iSemantic50169.2020.9234204>

Triwiniastuti, Chatarina Sitoresmi, Universitas Kristen, Satya Wacana, Wahyu Sabatini, Universitas Kristen, and Satya Wacana, 'Implementasi Program Ujian Nasional Di SMA Negeri', *Kelola: Jurnal Manajemen Pendidikan*, 20, 2019, 11–12

Walkington, Candace, 'Design Research on Personalized Probem Posing in Algebra', *Proceedings of the 39th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, 2017, 195–202

Weller, Joshua A., Nathan F. Dieckmann, Martin Tusler, C. K. Mertz, William J. Burns, and Ellen

Peters, 'Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach', *Journal of Behavioral Decision Making*, 26.2 (2013), 198–212 <https://doi.org/10.1002/bdm.1751>

Wijaya, T.T., Zhou Ying, Aditya Purnama, and Neni Hermita, 'Indonesian Students' Learning Attitude towards Online Learning during the Coronavirus Pandemic', *Psychology, Evaluation, and Technology in Educational Research*, 3.1 (2020), 17–25 <https://doi.org/10.33292/petier.v3i1.56>

Wijaya, T.T., Zhou Ying, and Lin Suan, 'Gender and Self-Regulated Learning During COVID-19 Pandemic in Indonesia', *Jurnal Basicedu*, 4.3 (2020), 725–32 <https://doi.org/10.31004/basicedu.v4i3.422>

Wind, Stefanie A., and Jessica D. Gale, 'Diagnostic Opportunities Using Rasch Measurement in the Context of a Misconceptions-Based Physical Science Assessment', *Science Education*, 99.4 (2015), 721–41 <https://doi.org/10.1002/sce.21172>

Wu, Xiaohong, Ying Zhou, and Zongzhao Mo, 'A Comparative Study on the Comprehensive Difficulty of Junior High School National Examination', *Journal On Education*, 2.4 (2020), 352–66

Xu, Sufen, and Yun Wang, 'Problems and Suggestions on Reform of Teacher Qualification Examination under the Background of National Unified Examination', in *Advances in Social Science, Education and Humanities Research*, 2018, CCXLVI, 314–17 <https://doi.org/10.2991/icpel-18.2018.74>

Yang, Der Ching, and Iwan Andi Sianturi, 'An Analysis of Singaporean versus Indonesian Textbooks Based on Trigonometry Content', *Eurasia Journal of Mathematics, Science and Technology Education*, 13.7 (2017), 3829–48 <https://doi.org/10.12973/eurasia.2017.00760a>

Yang, Shun Chin, Mei Yung Tsou, En Tzu Chen, Kwok Hon Chan, and Kuang Yi Chang, 'Statistical Item Analysis of the Examination in Anesthesiology for Medical Students Using the Rasch Model', *Journal of the Chinese Medical Association*, 74.3 (2011), 125–29 <https://doi.org/10.1016/j.jcma.2011.01.027>