ISSN 2502-5872 (Print)
ISSN 2622-3627 (Elektronik)

Volume 10 Number 3, August 2025, 721-738

DESIGN OF A MATHEMATICS TEST INSTRUMENT BASED ON PHARMACY VOCATIONAL TRAINING WITH RASCH MODEL ANALYSIS ON STUDENTS' MATHEMATICAL PROBLEM-SOLVING ABILITY

Hary Nurprio Utomo[1*], Muhtarom[2], Ida Dwijayanti[3]
[1,2,3]Master's Degree in Mathematics Education, Universitas PGRI Semarang, Central Java, Indonesia
*Correspondence: hary.august@gmail.com

ABSTRACT

This research aims to measure students' mathematical problem-solving abilities using a pharmacy-based vocational mathematics test instrument with Rasch model analysis. The development model in this research is 4D design by Thiagarajan, which includes the stages: define, design, development, and disseminate, but in this study, it is limited to the Development stage. Respondent data from 30 students with a total of 5 questions yielded results indicating that the expert validity of all five questions is very good. The reliability of test items and students was analyzed, revealing weak reliability for students and sufficient reliability for the questions; and the reliability of the interaction between individuals or students with the overall test items falls into the weak category. The overall difficulty level of the questions falls into the fairly good category, which includes items that are neither too difficult nor too easy, specifically in the moderate, easy, and difficult categories. In conclusion, the analysis of the pharmacy vocational-based mathematics test instrument using the Rasch model analysis is declared valid, reliable, and has a good level of difficulty.

Keywords: Rasch Model, Instrument, Mathematical Problem Solving

How to Cite: Utomo H. N., Muhtarom., & Dwijayanti, I. (2025). Design Of A Mathematics Test Instrument Based On Pharmacy Vocational Training With Rasch Model Analysis On Students' Mathematical Problem-Solving Ability. *Mathline: Jurnal Matematika dan Pendidikan Matematika*, *10*(3), 721-738. http://doi.org/10.31943/mathline.v10i3.867

PRELIMINARY

Evaluation is an important aspect of the learning process to identify the success of education (Herwin, 2019). According to Imam Darsono (2013), one of the efforts to improve the quality of education is the need for an enhancement in the quality of learning, namely through proper evaluation of both the process and the outcomes of learning. Evaluation is a process that needs to be carried out well to measure the actual abilities of the students themselves (Talakua, Ratumanan, and Tamalene, 2020). The evaluation process needs to be conducted properly to measure the learning process's achievement and ensure that students can receive the material well using appropriate

instruments (Isrotun, 2023). The appropriate instruments are those that have valid measuring power (Muluki & Bundu, 2020).

In developing test instruments, systematic steps are needed to achieve optimal results (Layne, Yli-Piipari, and Knox, 2021) so that the instruments are appropriate, meaning they can be used in high-quality assessments and are capable of measuring students' actual abilities (Stolt, Kottorp, and Suhonen, 2022). Test instruments are an important part of learning, as they can be used to evaluate and measure learners' actual abilities and as one way to assess someone's competence after studying something (Utomo, 2018). Therefore, the development of test instruments must be carried out carefully and professionally so that the results obtained and analyzed thoroughly reflect the student profile, which can serve as a basis for teachers in designing adaptive and responsive learning for student diversity (Elfitra, Armanto, and Dewi 2024).

Rasch can be an effective tool in the evaluation and selection of good questions, resulting in good and valid outcomes (Stolt, Kottorp, and Suhonen, 2022). The Rasch Model is also easy to use and apply, providing accurate analysis results (Che Lah, Tasir, and Jumaat, 2023), and it can review the likelihood of correctly answering questions by comparing students' abilities to the difficulty levels of the questions (Nielsen, Martínez-García, and Alastor, 2021). The developed questions can truly measure and determine the extent of students' abilities through test development and analysis of the items used (Hu, Jiang, and Bi, 2022).

Based on the survey results by researchers on vocational high school students in the pharmacy department in Pemalang indicate their lack of interest in studying mathematics. If this continues, it will affect their mathematical problem-solving abilities (Yuliati, 2021). This is in line. According to Sumandya and Widana, in mathematics education at vocational high schools, some students believe that mathematics has no connection to the vocation they are pursuing, resulting in very low interest in mathematics (Sumandya & Widana, 2022). According to (Sholihat & Dwiyanti, 2024)), it is necessary to integrate mathematical concepts into a broader context in studying mathematics. One of them is the need to connect the concepts of mathematics material with everyday life (Chotimah, Bernard, and Wulandari 2018). From this, teachers prepare vocational-based mathematics test instruments that analyze the quality of the questions in terms of their validity and reliability to effectively assess students' abilities because if the instruments have not been examined for item quality, it will affect the measurement of students abilities (Fauziana & Dessy Wulansari, 2021).

Therefore, In this study, the researcher is interested in creating a vocational-based test instrument specifically designed for vocational high school students, which, when tested, can accurately measure their mathematical abilities.

The test results obtained in the form of scores are then analyzed using the Rasch Model, which has advantages compared to classical test theory (Rahim & Haryanto, 2021). One of these advantages is that the probability of a subject answering an item correctly depends on the subject's ability and the item's characteristics (Adi et al., 2022). The Rasch Model, in analyzing the validity of an instrument, can be conducted from several aspects so the resulting instrument can be more reliable (Indihadi, Suryana, and Ahmad, 2022) and is considered better in terms of its consistency (Sumintono, 2018). Another advantage of Rasch modelling is the three reliabilities produced, namely person reliability, item reliability, and Cronbach's alpha. The Rasch model can predict missing data based on systematic response patterns (Aziz, 2015)**.**

In this study, the researcher uses the Rasch model approach to determine the quality of the test instrument used to assess students' mathematical problem-solving abilities on the topic of relations and functions. According to Muhtarom (2024), using Rasch modelling in instrument validation produces more holistic information about the instrument being developed. This quality is measured based on several indicators, namely the items that fit the Rasch model and the reliability of the items. Therefore, a test instrument was designed, and then it was determined which items fit and which did not fit the Rasch model. In addition, with the help of Winstep software, Cronbach's alpha value will be determined to assess the reliability of the test items.

METHODS

This research is a Research and Development (R&D) with a 4D model, namely Define, Design, Develop, and Disseminate (Sugiyono, 2019), limited to the Develop stage. At the define stage, the researcher identifies problems occurring in the classroom so that the content of the test instrument to identify mathematical problem-solving abilities can be adjusted and developed. This research was conducted over 3 weeks with respondents from the 11th-grade pharmacy program at SMK Al Manaar Muhammadiyah Pemalang with a total of 30 students selected for the material on relations and functions, starting from the Design stage where the researcher designed a draft of the mathematical problem-solving ability test instrument on relations and

functions. The development stage is where the test instrument is produced and validated by experts, and suggestions are received from expert validators and practitioners. The test instrument was then given to students, and the researcher analyzed the students' work results using the Rasch model analysis with the Winstep application. Rasch model analysis has the advantage of examining respondents' answers to each question and determining which respondents gave accurate or inaccurate answers for each item on the test. Rasch model analysis will determine how well a person can adapt to one or more factors, including item difficulty (Hissbach, Klusmann, and Hampe, 2011) and the likelihood that a person will give the right answer (Peeters & Stone, 2009).

The instrument used is vocational-based mathematics in the form of an essay to measure students' mathematical problem-solving abilities, consisting of 5 questions covering four aspects of problem-solving by Polya. The instrument is then given to subject matter experts, involving two experts from lecturers and two from mathematics teachers. The data from the assessment were then analyzed using descriptive techniques with percentages (Purwanto, 2010)by calculating the scores achieved from all aspects using the following formula:

$$N = \frac{\text{The evaluation results}'\text{score}}{\text{maximal score}} \times 100\%$$

Table 1. Expert Validation Result Criteria

Question item	Category
$86\% \leq N < 100\%$	Very good
$72\% \leq N < 85\%$	Good
$58\% \leq N < 71\%$	Fair
$44\% \leq N < 57\%$	Poor
$N \leq 44\%$	Very Poor

Data analysis

Using the Rasch model, the test results obtained are then analyzed using Winsteps software to determine the interaction between respondents and item statement elements and detect using logit values. These values are used to calculate scores as they can reflect the probability of item selection from a group of respondents (Wibisono, 2016) The data generated were analyzed based on instrument reliability, unidimensionality, item difficulty analysis, and item fit analysis.

Item Fit Order

In Rasch model analysis, item fit order is used to explain whether the instrument items typically function to perform measurement (Muhtarom, 2024). The Winsteps

software output obtained several item parameters that fit the Rasch model, along with Cronbach's alpha value, which is the result of the overall item reliability test. The criteria for examining the fit of items and persons that do not fit are based on Outfit MNSQ (means-square), Outfit ZSTD (Z-Standard), and point measure correlation (Pt Measure Corr) (Boone, 2014) as follows (Boone, 2014): Outfit MNSQ value (accepted if $0.5 < MNSQ < 1.5$), Outfit ZSTD value (accepted if $-2.0 < ZSTD < 2.0$) and Value $0.4 < Pt\ Measure\ Corr < 0.85$ (Sumintono, 2018).

Items that do not meet the three previously mentioned conditions are considered "not fit" and must be replaced; however, items that meet at least two of these requirements are still considered "fit" or in good condition (Sumintono, 2018).

Reliability

Reliability is a factor that affects the quality of an item, describing the stability of respondents and items (Duncan et al., 2003) with reliability values from the Rasch model ranging from zero to one, often known as Alpha Cronbach, with an acceptable range for the Alpha Cronbach coefficient of the instrument being 0.70 to 0.99 (Mohamad et al., 2015). Here is the person and item reliability table that serves as the basis for establishing the reliability of both people and items (Linacre, 2019).

Table 2. Person and Item Reliability

Value	Category
> 0.94	Excellent
0.91 – 0.94	Very good
0.81 – 0.90	Good
0.67 – 0.80	Fair
< 0.67	Poor

Level of item difficulty

Rasch's model divides item difficulty into four groups based on the item measurement (logit) and the standard deviation (SD) of the item logit (Boone, 2014; Sumintono, 2018).

Table 3. Item Difficulty Level Criteria

Measure (logit)	Interpretation
Measure (logit) $< -SD$	Easy
$-SD \leq$ Measure (logit) < 0.00	Medium
$0.00 \leq$ Measure (logit) $< SD$	Difficult
Measure (logit) $> SD$	Very difficult

Unidimensionality

The Rasch model's validity is assessed using the unidimensionality test. In this test, the evaluation of the interaction between the respondents' responses to the items and the responses to the items themselves is based on a single concealed characteristic (Ridho, 2011). The raw variance measure value, which should be at least 20%, is what is looked at in the unidimensionality test. However, if the instrument quality is less than 15%, it is considered good for cases of unexplained variance (Sumintono, 2018). The raw variance shows that the variable is unidimensional:

Table 4. Raw Variance Category

Value	Category
$< 20\%$	Very poor
$\geq 20\%$	Minimum
$20\% - 40\%$	Fair
$40\% - 60\%$	Good
$> 60\%$	Excellent

Person and Item Separation Index

Person and Item Separation Index functions to estimate a tool that can differentiate students' skills. The wider the spread of items from easy to difficult, and the higher the person and item separation index, the more accurately the item distribution responds to the items (Linacre, 2019). There is a value for the separation index that ranges from 0 to infinity; the higher the separation, the better. Table 10 displays the criteria for the person and item separation indices (Linacre, 2019; Boone. 2014).

Table 5. Person and Item Separation Index Criteria

Value	Category
> 5	Excellent
$3 - 4$	Very good
$2 - 3$	Good
≥ 1.5	Acceptable
< 1.5	Not acceptable

RESULT AND DISCUSSION

Stage 1: Define

In the define stage, the initial step taken is to involve and clearly define the learning objectives measured by the created questions, including identifying the initial problem-solving abilities of the students. The formulation of mathematics questions in this research is vocational-based using the mathematical problem-solving steps by

George Polya, where according to (Sepriyanti, Trinova, and Susanto, 2020)The use of Polya's steps can help students solve problems. By implementing Polya's four steps, students are guided in understanding a problem until they conclude the results of their work. Problem-solving using Polya's steps can be organized systematically and practically, making it easier for students (Arifin and Aprisal, 2020).

There are 2 cognitive domains included in the preparation of HOTS question items, namely C3 (applying) and C4 (analyzing). Here are the indicators for the research on the development of problem-solving instruments:

Table 6. Research Indicators for Instrument Development

No	Indicator	Assessment Rubric
1	Understanding the Problem	Writing down the known and asked information correctly and completely and Elaborating the answer according to the question
2	Devising a plan	Preparation of a resolution plan accurately and systematically and Connecting with information from the given questions accordingly
3	Carrying out the plan	Perform calculations accurately and systematically and Connect with information from the given question accordingly
4	Look Back	Make a conclusion based on the concept used in verifying its truth.

Stage 2: Design

In the design stage, the preparation of question items based on the cognitive domain involves creating 5 vocational-based mathematics essay questions, which are then validated by 4 experts, namely 2 lecturers and 2 mathematics teachers at the school, as the instrument question validators. The type of pharmacy vocational-based test instrument created is :

Question number 1

In question number 1, the given question is related to determining the number of relations from the domain to the codomain with a HOTS C3 level. The generated question is as follows:

Given a pharmaceutical company has three different types of drug products, namely Drug A, Drug B, and Drug C, each of which must be packaged in a unique dosage form and there must be no repetition of dosage forms for different drugs. The company has four available main dosage form options: Capsules, Tablets, and Syrups. If each type of medicine must be paired sequentially with one of the available forms of preparation without any form of preparation being used more than once, how many different ways are there to package the three types of medicine with the available forms of preparation?

Question number 2

In question number 2, the given question is related to determining the value of a function at the HOTS C3 level. The resulting question is as follows:

In a primary healthcare facility in a remote area, a medical professional is reviewing the medication administration protocol for pediatric patients. One of the essential medications available for treating certain conditions has an established adult dosage standard of 400 mg. To ensure the safety and effectiveness of therapy in children, the medical staff consistently refers to the 'Young's Rule' principle in determining pediatric dosages, where its formula is stated as:

$$\text{child's dose} = \left(\frac{\text{child}'\text{age}}{\text{child}'\text{age}+12}\right) \times \text{adult dose}$$

In a certain case, a pediatric patient presented with symptoms requiring the intervention of that medication, and after a thorough clinical assessment, it was decided that the child would receive an individual dose of 300 mg to achieve optimal therapeutic effects without causing toxicity. Based on this scenario and the precise application of Young's Rule, what is the estimated age of the pediatric patient in years?

Question number 3

In question number 3, the given question is the same as question number 2 but at a higher level, where the requested function is not yet known. Therefore, to obtain the result, it is necessary to create a model of the function at HOTS C4 level. The resulting question is as follows:

A leading pharmaceutical company, "Herba Farma," is in the strategic planning stage to launch its latest innovation, a revolutionary drug named "NeuroBoost," which is expected to enhance cognitive function. In an effort to optimize the production process and ensure financial sustainability, the management team needs to understand the cost structure in depth. After conducting an initial analysis, it was found that the fixed costs associated with the operation of the production facility, such as factory rent, equipment depreciation, and administrative staff salaries, amount to Rp 100,000,000 per month, regardless of the production volume. Meanwhile, the variable cost per unit of NeuroBoost production, which includes active raw materials, primary and secondary packaging, and direct labor wages, is set at Rp 5,000 per unit of NeuroBoost produced. Currently, the management of Herba Farma is considering an initial production target of 50 units of NeuroBoost for a small-scale market test. Considering all the cost

components, calculate in detail the total production cost that Herba Farma will incur if they decide to produce 50 units of NeuroBoost?

Question number 4

In question number 4, the given question is the same as question number 3 but at a higher level where the requested function is not yet known, so to obtain the result, it is necessary to create a model of the function at the HOTS C4 level. The resulting question is as follows:

In clinical pharmacy practice, determining the correct drug dosage for pediatric patients is a crucial aspect that requires precision and a deep understanding of various calculation methods. One of the methods known for its accuracy in considering the differences in drug metabolism in children based on their body size is the Lund-Powell-Talbot (LPT) method, as reported in the British Journal of Anaesthesia.

The LPT method outlines two specific dosage calculation schemes, depending on the child's weight category:

- For children weighing less than 30 kg, the recommended dosage is a percentage of the adult dosage calculated by multiplying the child's weight by a factor of two.
- Meanwhile, for children weighing more than 30 kg, the percentage of the adult dosage is determined by adding 30 to the child's weight.

A pediatrician prescribed medication to a young patient who at that time weighed less than 30 kg, with a single dose of 150 mg. Considering that the standard dose for adult patients of the same medication is 600 mg, analyze and accurately determine the exact weight of the child. Next, imagine a pharmacist has to prepare a dose of medication for another child patient weighing 60 kg; using the principles of the LPT method, calculate the single dose of medication that the pharmacist should administer to that patient.

Question number 5

In question number 5, the question given is related to determining the value of a function at the HOTS C3 level as in question number 2. The resulting question is as follows:

In a busy pediatric clinic, Dr. Hary is reviewing the medication prescription for his young patient, Budi, who is 8 years old and weighs 25 kg. Based on the hospital's standard protocol for handling certain infections, the single dose of the 'Vitasol A' medication for children is determined using the formula $D_{single} = 0.025W$, where D_{single} is the dose in milligrams (mg) to be given in a single intake and W represents the child's

weight in kilograms (kg). Budi is scheduled to take this medication twice a day, every morning and evening, to ensure the effectiveness of continuous treatment. The clinic pharmacy currently only provides 'Vitasol A' in a single bottle with a total capacity of 125 mg. Considering Budi's daily dosage needs and the frequency of medication administration, how many days will one bottle of 'Vitasol A' be sufficient to meet all of Budi's medication needs?

And here are the results of the validation by four validators with validation criteria in terms of material coverage, question presentation construction, problem-solving indicators, and the language used. From the five items above, the result is that they are considered suitable for use, and here are the validation results shown in the following table:

Table 7. Expert Validation Results

Item	Validator				Average	Category
	I	II	III	IV		
Item 1	85%	84%	88%	90%	87%	Very good
Item 2	87%	86%	87%	92%	88%	Very good
Item 3	86%	85%	88%	90%	87%	Very good
Item 4	90%	87%	88%	93%	89%	Very good
Item 5	92%	87%	88%	95%	91%	Very good

Based on the data presented above, the five items show very high consistency in the assessment by the four validators. The main points of the validation results include the presence of Consistent Quality, where all items indicate that each item is well-designed and meets the expected validity standards. This indicates that the validated materials or instruments have consistent and high quality throughout all their parts.

Stage 3: Development

The questions that experts have validated are then tested on a small scale, where these questions are given to 11th-grade students of SMK Al Manaar Muhammadiyah Pemalang. The test results are then presented in the form of Microsoft Excel and analyzed using the Winstep program. Information on the analysis results using the Rasch Model includes:

Instrument Validity Analysis

The results of the item validity analysis using the Rasch model are considered very effective because they yield reliable validity analysis results (Sari et al., 2016).

Here are the results of the content validity analysis using the Winstep program, which yielded output results in the form of Misfit Order, as shown in the table below :

Table 8. Output Misfit Order

Outfit		PT-Measure		Item
MNSQ	ZSTD	CORR.	EXP.	
1.21	0.9	0.57	0.56	Question 5
1.14	0.6	0.55	0.58	Question 2
1.10	0.5	0.50	0.58	Question 4
0.83	-0,7	0.60	0.57	Question 3
0.74	-1.1	0.66	0.57	Question 1

From the results of table 8 above, then each item is analyzed to determine the validity criteria of each item as shown in the following table :

Table 9. Item Analysis Results

No	Item Number	Measurement Accuracy Criteria			Decision
		Outfit MNSQ	Outfit ZSTD	PT-Meancorr	
1	Question 5	1.21	0.9	0.57	Valid
2	Question 2	1.14	0.6	0.55	Valid
3	Question 4	1.10	0.5	0.50	Valid
4	Question 3	0.83	-0.7	0.60	Valid
5	Question 1	0.74	-1.1	0.66	Valid

Based on the analysis results of the table above, viewed from the validity of the test items using the Rasch model, it was found that 5 items fall into the valid category because they meet all categories of MNSQ, Outfit ZSTD, and Point Measure Correlation. (Pt measure Corr). From these results, the questions have met the criteria and can ensure that the students' problem-solving abilities are indeed tested through appropriate and high-quality items.

Instrument Reliability Analysis

Test's reliability is determined by how consistently its results are achieved (Langenfeld et al., 2020). When the same conditions are repeated, test items with a high dependability rating yield consistent and reliable findings (Hasnida and Ghazali 2016). based on the analysis findings of the study's instruments, as shown in the following Summary Statistics table.

Table 10. Summary Statistic : Measurement of Persons

Mean	Standard Deviation	Max	Min	Separation	Person Reliability
11.9	2.4	16.0	7.0	0.92	0.46

Table 11. Summary Statistic : Measurement of Item Responses

Mean	Standard of Deviation	Max	Min	Separation	Item Reliability
71.4	8.2	81.0	57.0	1.83	0.77

The analysis of the summary statistic output on Winstep found that the data output is in the form of a table divided into 2 types. The first table contains a summary of the measurement of persons or students, and the second is a summary of the item responses based on the consistency of the 5 items. The reliability score for students is 0.46, which means the students' reliability is weak because the score is less than 0.67. Meanwhile, the reliability score for the questions is 0.77, which means the reliability of the questions is sufficient, falling within the range of 0.67 – 0.80. Reliability is the interaction between the person or student and the overall items of the questions, with a Cronbach's alpha value of 0.49, indicating that the students' reliability falls into the weak category because the score is less than 0.67. This assessment is based on the Cronbach's alpha value, which can measure the interaction between the person and the overall items of the questions. Thus, it can be concluded that the level of consistency of students in taking the test is weak.

Person and Item Separation Index

The grouping of persons and items can be determined through the separation value where the larger the separation value, the better the overall quality of the instrument in identifying respondent groups and item groups. For a Person separation value of 0.92, $H = \frac{[(4\times0.92)+1]}{3} = 1.56$ rounded to 2, this indicates that there are two groups of respondents which can be interpreted as respondents with high ability and respondents with low ability. And for a Person separation value of 1.83, $H = \frac{[(4\times1.83)+1]}{3} = 2.77$ rounded to 3, this indicates that the item is of good quality because it divides the items into 3 groups, namely high, medium, and low. Thus, it can be concluded that the discriminative power is already capable of distinguishing the problem-solving abilities of students with high and low capabilities.

Unidimensionality Analysis

Unidimensionality of an instrument is a measure to evaluate the analysis of the developed instrument related to its ability to measure what it is supposed to measure.

Table 12. Unidimensionality

Residual Variance	Empirical	Modeled
Total Raw Variance in Observations	100%	100%
Raw Variance Explained by Measure	39.9%	39.3%
Raw Variance Explained by Persons	17.2%	16.9%
Raw Variance Explained by Item	22.7%	22.4%

From the above Figure 3, the construct validity results show that the Raw variance explained by the measure, as explained by empirical measurement, reached a value of 39.9%, while in the Rasch model it is predicted to be 39.3%. This indicates that the unidimensionality requirement can be met, so the developed instrument is considered capable of measuring what it is supposed to measure, namely the ability to solve mathematical problems with the material of relations and functions.

Analysis of Item Difficulty Level

The difficulty level of an item determines how many respondents are likely to answer the item correctly. In Rasch modeling, to see how difficult an item is based on the results of the Item Measure table. The logit value for each item is displayed in the Item Measure, which is sorted from highest to lowest. The logit value is a measure of item difficulty. The more difficult the test item, the higher its logit value. A question can be considered good if its difficulty level is proportional (balanced) (Rusiyah, Eraku, and Supadmi, 2020).

Table 13. Item Statistic : Measure Order

Entry Number	Total Score	Total Count	Measure
5	57	30	1.00
2	69	30	0.16
4	73	30	-0.11
3	77	30	0.39
1	81	30	-0.66

From the results in table 13, the distribution of the difficulty levels of the test items is categorized into four categories as follows (Erfan et al. 2020).

Table 14. Item Difficulty Levels

No. Item	Logit value	Interpretation
1	-0,66	Easy
2	0,16	Difficult
3	-0,39	Medium
4	-0,11	Difficult
5	1	Very difficult

The results of the difficulty level analysis of the pharmacy vocational-based mathematics test items can be interpreted based on the measure logit values, which describe the relative difficulty level of each item in the test. In the table above, the items are dominated by the difficult category with 2 out of 5 items (40%), 1 item in the moderate category (20%), 1 item in the easy category (20%), and 1 item in the very easy category (20%). According to Nuryanti et al (2018) an item is accepted if it is neither too difficult nor too easy, in other words, the difficulty level of the item falls into the moderate, easy, and difficult categories.

Rasch modeling can provide detailed information about the difficulty level of the given instrument (Curtis and Boman 2007). Furthermore, Ismail et al (2010) stated that through Rasch modeling, a good test instrument can be obtained to measure respondents' understanding levels in research, making its use highly recommended, especially in the field of education.

CONCLUSION

Analysis of the pharmacy vocational-based mathematics test instrument on problem-solving ability using the Rasch model on relations and functions material for class XI at SMK Al Manaar Muhammadiyah Pemalang, obtained a question validity level of 5 questions. Then, from the students' reliability scores, the analysis showed that a student might answer some questions correctly but then make random mistakes on other questions that they should be able to answer, or vice versa. From the reliability scores of the questions, the results indicated that the questions functioned quite well in distinguishing between students with different abilities, and the reliability between individuals or students with the overall items fell into the category of being inconsistent in measuring the same construct, thus necessitating a retest to yield different results. The overall difficulty level of the questions falls into the fairly good category, which includes items that are neither too difficult nor too easy, particularly in the moderate, easy, and difficult categories.

The recommendation in this study is for more in-depth research on the development of vocational-based pharmacy test instruments on students' problem-solving abilities, both from researchers and teachers, to obtain more accurate results regarding the quality of the instruments and the improvements needed to ensure that the test instruments used are truly valid and reliable. This study also has limitations, as the development of questions is only within the scope of the pharmacy vocational field, so

further research with other vocational fields is needed to ensure that these results can contribute to the development of better vocational-based assessment instruments in the future.

ACKNOWLEDGMENT

The authors would like to express their appreciation to DPPM Kemdiktisaintek for facilitating research funding assistance, which enabled the smooth execution of this research. The authors would like to express their appreciation to the Mathematics Education Postgraduate Program, PGRI University Semarang for granting the research permission. And Also we would like to thank the XI Pharmacy students at SMK Al Manaar Muhammadiyah Pemalang who have been the data source for this research.

REFERENCES

Adi, Nur Romdlon M, Hidar Amaruddin, Habis Maulana M Adi, and Isna Laili Qurroti A'yun. 2022. "Journal of Educational Research and Evaluation Validity and Reliability Analysis Using the Rasch Model to Measure the Quality of Mathematics Test Items of Vocational High Schools." *Journal of Educatinal Research and Evaluation* 11(1): 103–13. http://journal.unnes.ac.id/sju/index.php/jere.

Arifin, Sartika, and Aprisal Aprisal. 2020. "Penerapan Model Pembelajaran Kooperatif Tipe Pair Checks Terhadap Kemampuan Pemecahan Masalah Matematika." *Jurnal Pendidikan Matematika* 11(1): 89. doi:10.36709/jpm.v11i1.9974.

Aziz, Rahmat. 2015. "Aplikasi Model Rasch Dalam Pengujian Alat Ukur Kesehatan Mental Di Tempat Kerja." *Jurnal Psikologi Islam (JPI)* 2. doi:https://doi.org/10.18860/PSI.V12I2.6402.

Boone, W. J, Staver, R. J dan Yale, S. M. 2014. *Rasch Analysis in the Human Sciences*. London: Springer. doi:http://dx.doi.org/10.1007/978-94-007-6857-4.

Budi Utomo. 2018. "Analisis Validitas Isi Butir Soal Sebagai Salah Satu Upaya Peningkatan Kaulitas Pembelajaran Di Madeasah Berbasis Nilai-Nilai Islam." *Jurnal Pendidikan Matematika* 1. doi:http://dx.doi.org/10.21043/jmtk.v1i2.4868.

Che Lah, Noor Hidayah, Zaidatun Tasir, and Nurul Farhana Jumaat. 2023. "Applying Alternative Method to Evaluate Online Problem-Solving Skill Inventory (OPSI) Using Rasch Model Analysis." *Educational Studies* 49(4): 644–66. doi:10.1080/03055698.2021.1874310.

Chotimah, Siti, M. Bernard, and S. M. Wulandari. 2018. "Contextual Approach Using VBA Learning Media to Improve Students' Mathematical Displacement and Disposition Ability." In *Journal of Physics: Conference Series*, Institute of Physics Publishing. doi:10.1088/1742-6596/948/1/012025.

Curtis, David D, and Peter Boman. 2007. "X-Ray Your Data with Rasch." *International Education Journal* 8(2): 249–59. http://iej.com.au.

Duncan, Pamela W., Rita K. Bode, Sue Min Lai, and Subashan Perera. 2003. "Rasch Analysis of a New Stroke-Specific Outcome Scale: The Stroke Impact Scale." *Archives of Physical Medicine and Rehabilitation* 84(7): 950–63. doi:10.1016/S0003-9993(03)00035-2.

Elfitra, Elfitra, Dian Armanto, and Izwita Dewi. 2024. "Assessing Student Characteristics With The Right Instruments: An Important Pathway To Implementing A Merdeka Curriculum In Indonesia." *Mathline : Jurnal Matematika dan Pendidikan Matematika* 9(4): 1153–68. doi:10.31943/mathline.v9i4.703.

Erfan, Muhammad, Mohammad Archi Maulyda, Vivi Rachmatul Hidayati, Fitri Puji Astria, and Tursina Ratu. 2020. "Analisis Kualitas Soal Kemampuan Membedakan Rangkaian Seri Dan Paralel Melalui Teori Tes Klasik Dan Model Rasch." *Indonesian Journal of Educational Research and Review* 3(1): 11. doi:https://doi.org/10.23887/ijerr.v3i1.24080.

Fauziana, Anis, and Andhita Dessy Wulansari. 2021. "Analisis Kualitas Butir Soal Ulangan Harian Di Sekolah Dasar Dengan Model Rasch." *Jurnal Kependidikan Dasar Islam Berbasis Sains* 6. doi:https://doi.org/10.21154/ibriez.v6i1.112.

Hasnida, Nor, and Md Ghazali. 2016. "A Reliability and Validity of an Instrument to Evaluate the School-Based Assessment System: A Pilot Study." *International Journal of Evaluation and Research in Education (IJERE)* 5(2): 148–57. doi:http://doi.org/10.11591/ijere.v5i2.4533.

Herwin. 2019. "Evaluation of Social Studies Learning Program at Sekolah Dasar Negeri 126 Lagoe." *DIDAKTIKA: Jurnal Pendidikan Sekolah Dasar* 2: 41–48. https://journal.uny.ac.id/index.php/didaktika.

Hissbach, Johanna C., Dietrich Klusmann, and Wolfgang Hampe. 2011. "Dimensionality and Predictive Validity of the HAM-Nat, a Test of Natural Sciences for Medical School Admission." *BMC Medical Education* 11(1). doi:10.1186/1472-6920-11-83.

Hu, Xinyang, Yanxia Jiang, and Hualin Bi. 2022. "Measuring Science Self-Efficacy with a Focus on the Perceived Competence Dimension: Using Mixed Methods to Develop an Instrument and Explore Changes through Cross-Sectional and Longitudinal Analyses in High School." *International Journal of STEM Education* 9(1). doi:10.1186/s40594-022-00363-x.

Imam Darsono. 2013. *Peran Guru Dalam Proses Pembelajaran*.

Indihadi, Dian, Dodi Suryana, and Aslina Binti Ahmad. 2022. "The Analysis of Construct Validity of Indonesian Creativity Scale Using Rasch Model." *Creativity Studies* 15(2): 560–76. doi:10.3846/cs.2022.15182.

Ismail, Issham, Rozhan Mohammed Idrus, and Siti Sarah Mohd Johari. 2010. "Acceptance on Mobile Learning via SMS: A Rasch Model Analysis." *International Journal of Interactive Mobile Technologies (iJIM)* 4(2): 10. doi:10.3991/ijim.v4i2.1144.

Isrotun, Umi. 2023. "Analisis Kualitas Instrumen Untuk Mengukur Kreatifitas Siswa Melalui Pembelajaran Berdiferensiasi." *Jurnal Pendidikan Berkarakter* 1(4). doi:10.51903/pendekar.v1i4.528.

Langenfeld, Thomas, Jay Thomas, Rongchun Zhu, and Carrie A. Morris. 2020. "Integrating Multiple Sources of Validity Evidence for an Assessment-Based Cognitive Model." *Journal of Educational Measurement* 57(2): 159–84. doi:10.1111/jedm.12245.

Layne, Todd, Sami Yli-Piipari, and Tony Knox. 2021. "Physical Activity Break Program to Improve Elementary Students' Executive Function and Mathematics Performance." *Education 3-13* 49(5): 583–91. doi:10.1080/03004279.2020.1746820.

Linacre, John M.. 2019. *A User's Guide to WINSTEPS Ministep : Rasch-Model Computer Programs*. Winsteps.com. https://www.researchgate.net/publication/238169941_A_User''s_Guide_to_Winsteps_Rasch-Model_Computer_Program (May 3, 2025).

Mohamad, Mimi Mohaffyza, Nor Lisa Sulaiman, Lai Chee Sern, and Kahirol Mohd Salleh. 2015. "Measuring the Validity and Reliability of Research Instruments." *Procedia - Social and Behavioral Sciences* 204: 164–71. doi:10.1016/j.sbspro.2015.08.129.

Muhtarom, Muhtarom. 2024. "Developing an Instruments to Measure Prospective Teacher Beliefs about Mathematical Problem-Solving Using the Rasch Model." *Jurnal Elemen* 10(2): 274–88. doi:10.29408/jel.v10i2.25040.

Muluki, Ardilah, and Patta Bundu. 2020. "Analisis Kualitas Butir Tes Semester Ganjil Mata Pelajaran IPA Kelas IV Mi Radhiatul Adawiyah." *Jurnal Ilmiah Sekolah Dasar* 4(1): 86–96. doi:https://doi.org/10.23887/jisd.v4i1.23335.

Nielsen, Tine, Inmaculada Martínez-García, and Enrique Alastor. 2021. "Critical Thinking of Psychology Students: A within- and Cross-Cultural Study Using Rasch Models." *Scandinavian Journal of Psychology* 62(3): 426–35. doi:10.1111/sjop.12714.

Nuryanti, Sri, Muhammad Masykuri, and E. Susilowati. 2018. "Analisis Iteman Dan Model Rasch Pada Pengembangan Instrumen Kemampuan Berpikir Kritis Peserta Didik Sekolah Menengah Kejuruan." *Jurnal Inovasi Pendidikan IPA* 4(2): 224–33. doi:10.21831/jipi.v4i2.21442.

Peeters, Michael J, and Gregory E Stone. 2009. "An Instrument to Objectively Measure Pharmacist Professionalism as an Outcome: A Pilot Study." *The Canadian Journal of Hospital Pharmacy* 3. doi:https://doi.org/10.4212/cjhp.v62i3.790.

Purwanto. 2010. *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Belajar.

Rahim, Abdul, and Haryanto Haryanto. 2021. "Journal of Educational Research and Evaluation Implementation of Item Response Theory (IRT) Rasch Model in Quality Analysis of Final Exam Tests in Mathematics Article Info." *Journal of Educational Research and Evaluation* 10(2): 57–65. http://journal.unnes.ac.id/sju/index.php/jere.

Ridho, Ali. 2011. "Multidimensionalitas Pada Tes Potensi Akademik." *Presented at The Second International Conference of Indigenous and Cultural Psychology*: 1–16. http://repository.uin-malang.ac.id/1855/ (January 11, 2025).

Rusiyah, Sunarty S Eraku, and Supadmi. 2020. "Analisis Soal Ujian Akhir Semester Mata Pelajaran Geografi Dengan Menggunakan Permodelan Rasch." *Jurnal Swarnabhumi : Jurnal Geografi dan Pembelajaran Geografi* 5. doi:https://doi.org/10.31851/swarnabhumi.v5i1.4136.

Sari, Dini Riyantini, Nanan Sekarwana, Zahrotur Rusyda Hinduan, and Bambang Sumintono. 2016. "Analisis Tingkat Kepuasan Masyarakat Terhadap Dimensi Kualitas Pelayanan Tenaga Pelaksana Eliminasi Menggunakan Pemodelan Rasch." *Jurnal Sistem Kesehatan* 2(1). doi:10.24198/jsk.v2i1.10419.

Sepriyanti, Nana, Zulvia Trinova, and Andi Susanto. 2020. "The Application of The Polya's Steps Reviewed from Problem-Solving Ability in Two-Variable Linear Equation System (SPLDV)." *Tarbiyah : Jurnal Ilmiah Kependidikan* 9(1): 51. doi:10.18592/tarbiyah.v9i1.3543.

Sholihat, M. Nuur'aini, and Widya Dwiyanti. 2024. "Students' Interest in Learning Mathematics through Aquascape Project-Based Learning at Vocational High School." *Mathline : Jurnal Matematika dan Pendidikan Matematika* 9(4). doi:10.31943/mathline.v9i4.585.

Stolt, Minna, Anders Kottorp, and Riitta Suhonen. 2022. "The Use and Quality of Reporting of Rasch Analysis in Nursing Research: A Methodological Scoping Review." *International Journal of Nursing Studies* 132. doi:10.1016/j.ijnurstu.2022.104244.

Sugiyono. 2019. *Metode Penelitian Kuantitatif, Kualitatif Dan R&D*. Bandung: Alfabeta.

Sumandya, Wayan, and Wayan Widana. 2022. "Reconstruction of Vocational-Based Mathematics Teaching Materials Using a Smartphone." *Journal of Education Technology* 6(1): 133–39. doi:10.23887/jet.v6i1.4.

Sumintono, B. 2018. *Rasch Model Measurements as Tools in Assessment for Learning*. doi:https://doi.org/10.2991/icei-17.2018.11.

Talakua, Marlyd, Tanwey Gerson Ratumanan, and Hanisa Tamalene. 2020. "Komparasi Hasil Belajar Siswa Yang Menggunakan Media Swishmax Dan Tanpa Swishmax Pada Materi Operasi Hitung Pecahan Di Kelas VII SMP N 3 Ambon." *BAREKENG: Jurnal Ilmu Matematika dan Terapan* 14(1): 031–038. doi:10.30598/barekengvol14iss1pp031-038.

Wibisono, Susilo. 2016. "Aplikasi Model Rasch Untuk Validasi Instrumen Pengukuran Fundamentalisme Agama Bagi Responden Muslim." *Jurnal Pengukuran Psikologi dan Pendidikan Indonesia* 5(1). doi:http://dx.doi.org/10.15408/jp3i.v5i1.9239.

Yuliati, Ikha. 2021. "Kemampuan Pemecahan Masalah Matematika Ditinjau Dari Minat Belajar Peserta Didik." 05(02): 1159–68. *Jurnal Cendekia: Jurnal Pendidikan Matematika*. https://doi.org/10.31004/cendekia.v5i2.547