Jurnal Elemen, 12(2), 537-553, April 2026
https://doi.org/10.29408/jel.v12i2.33850

e- ISSN: 2442 - 4226

Evaluation of pharmacy mathematics assessment items using the Rasch model

Alifa Sabrina [1] *, Fitri Savitri [1], Farida Tuahuns [2]

[1] Pharmacy Associate Degree Program, Sekolah Tinggi Ilmu Kesehatan IKIFA, Jakarta, Indonesia
[2] Bachelor of Pharmacy Program, Sekolah Tinggi Ilmu Kesehatan IKIFA, Jakarta, Indonesia

* Correspondence: alifasabrina@ikifa.ac.id

Abstract

Evaluation of the quality of assessment instruments is an important factor in pharmacy mathematics learning because it has a direct impact on the accuracy of academic decision-making. This study aims to evaluate pharmacy mathematics assessment items by utilising the Rasch model approach to analyse the items. The methodology used was a quantitative design involving 403 students from the pharmacy study programme, consisting of 88 students in the regular programme and 315 students in the non-regular programme. The instrument analysed was a pharmacy mathematics test consisting of 15 questions. Data analysis included estimating item difficulty levels, model fit (INFIT and OUTFIT), and examining the measurement structure. The results of the analysis indicated that all items met the Rasch Model fit criteria, with item difficulty levels ranging from −1.50 to 1.75 logits. This instrument effectively assessed students' abilities across a range, from low to high levels, although it still had limitations in measurement accuracy at very high ability levels. These findings indicate that the measuring instrument has sufficient validity and reliability to be applied in the context of pharmacy mathematics learning. The implications of this study emphasise the importance of further developing the instrument to improve accuracy and fairness in measuring students' abilities.

Keywords: assessment instrument; mathematics learning outcomes; Rasch model analysis

How to cite: Sabrina, A., Savitri, F., & Tuahuns, F. (2026). Evaluation of pharmacy mathematics assessment items using the Rasch model. *Jurnal Elemen, 12*(2), 537-553. https://doi.org/10.29408/jel.v12i2.33850

Received: 6 January 2026 | Revised: 20 January 2026
Accepted: 3 April 2026 | Published: 30 April 2026

Introduction

Mathematics is not the only discipline that is the main focus in the context of pharmaceutical science, but it plays a very important role in various aspects related to mathematical calculations in pharmacy. The application of mathematical principles in the formulation and combination of pharmaceutical preparations, the determination of the ideal dosage based on individual body weight, and pharmacokinetic and pharmacodynamic modeling to predict drug concentrations in the body are vital (Derendorf & Schmidt, 2020; Wagner, 2020). Errors in calculating active ingredient concentrations or drug dosages can result in suboptimal effects or even endanger patient safety. An analysis conducted by Mulac et al (2022) revealed 100 incidents of numerical errors in drug dosage calculations, 78 of which had negative consequences for patients. In addition, a report by Haute (2021) stated that the most common medication errors, namely 80 out of 169 cases of incorrect dosages, were caused by calculation errors.

The difficulties faced by students in understanding lecture material, especially in mathematics and science, stem from a lack of mastery of basic concepts that should have been learned during secondary school education (Salafiya et al., 2025). Mastery of basic concepts in mathematics is an important element that students must have in order to follow more complex material in higher education. Research by Dwijayanti et al (2021) highlights the importance of understanding algebraic concepts for students, as errors in understanding basic concepts can hinder the application of more advanced mathematical calculations.

Previous research in the field of Pharmaceutical Mathematics and related disciplines generally analyzes the quality of assessment instruments using the Classical Test Theory (CTT) approach, including reliability analysis using Cronbach's alpha and content validity based on expert assessment (Hambleton & Swaminathan, 1985). Research in the field of pharmaceutical education and health sciences shows that CTT remains the most widely used approach due to its ease of application, despite its significant shortcomings. These shortcomings include dependence on item parameters related to the sample, limitations in the interpretation of raw scores, and the inability to evaluate the fit between individual abilities and items simultaneously (Crocker & Algina, 2008). In addition, the application of the parametric Rasch model in mathematics assessment usually requires a large sample size and certain distribution assumptions that are often difficult to fulfill, especially in the context of Pharmacy Mathematics assessment, which tends to be contextual and applicative (Embretson & Reise, 2000).

Although these approaches have made significant contributions to instrument assessment, previous studies have not consistently provided an evaluation framework that can combine item characteristics, participant abilities, and measurement accuracy in a single fixed interval scale. Therefore, the Rasch model was chosen in this study because of its advantages in producing measurements that are not influenced by samples or tests, as well as its ability to detect problems in items such as misfit, inaccuracy of difficulty levels, and weaknesses in targeting student abilities (Boone, 2016). If alternative assessment approaches such as CTT are used, there will be limitations in understanding the diagnostic quality of items and reducing the accuracy of measurements for assessment decisions (Bond & Fox, 2007)

Previous studies have indicated that the use of Rasch models in the validity and reliability aspects of measurement tools makes an important contribution to improving the quality, transparency, and accountability of assessment in the health education sector. However, the application of this method in the evaluation of Pharmacy Mathematics questions is still rarely reported, creating opportunities for research oriented towards construct-based measurement validity in this context (Cook & Beckman, 2006; De Champlain, 2010).

Based on these gaps, this study has a specific objective to evaluate the characteristics of questions in the Pharmacy Mathematics examination using the Rasch Model. The focus of the evaluation is on analysing the estimation of the difficulty level of each question item, the reliability of the instrument, and the dimensional structure of the instrument. Through this approach, the study is expected to provide more accurate and comprehensive evidence of measurement validity, as well as contribute to the development of a more precise and fair assessment system in Pharmacy Mathematics learning.

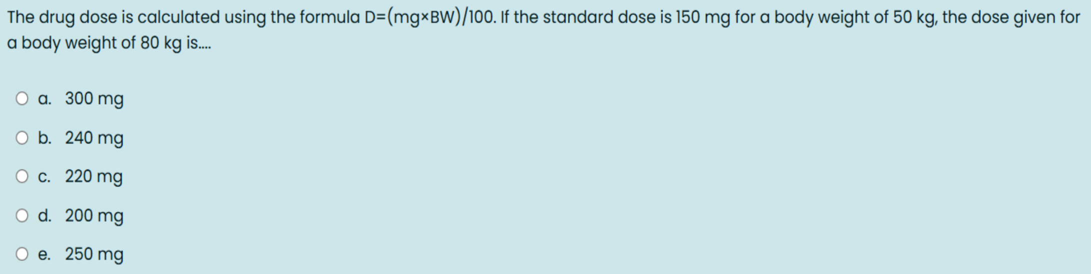
Methods

This study used a quantitative research design to evaluate the characteristics of assessment instruments aimed at measuring the mathematical abilities of pharmacy students (Sugiyono, 2014). This instrument was designed to assess students' abilities in performing pharmaceutical calculations. Purposive sampling technique was applied to select students who could be used as samples in this study, based on the material that students had acquired in pharmaceutical mathematics learning. The reason behind this purposive selection was that these students were more likely to be familiar with abstract reasoning in the context of pharmacy The participants in this study consisted of 403 pharmacy students who had taken the Pharmacy Mathematics course, with the following details:

Table 1. Demographic characteristics sample

Academic Program	Gender		Total
	Male	Female	
Reguler	29	59	88
Non-Reguler	42	273	315
Total	71	332	403

Data collection was conducted using a multiple-choice test consisting of 15 contextual mathematics questions in the field of pharmacy which aimed to measure pharmaceutical calculation skills using mathematical concepts. with the following example questions:



The drug dose is calculated using the formula D=(mg×BW)/100. If the standard dose is 150 mg for a body weight of 50 kg, the dose given for a body weight of 80 kg is....

○ a. 300 mg
○ b. 240 mg
○ c. 220 mg
○ d. 200 mg
○ e. 250 mg

Figure 1. Example question

Data analysis was performed using the Rasch Model in JMetrik software (statistical fit analysis, reliability, item-person map). The JMetrik software is utilised in Rasch Model analysis because it provides in-depth, accurate, and relevant psychometric analysis facilities for examining item characteristics in an educational context using Rasch Model (Prasetya, 2023). The unidimensionality assumption test used principal components analysis of the residuals. Item fit evaluation included MNSQ infit and outfit, with the following condition:

Table 2. Rasch model criteria for validity, reliability, and measurement fit

Aspect Evaluated	Indicator	Acceptable Criteria	Interpretation	Key References
Construct Validity	Item Fit (Infit & Outfit MNSQ)	0.5 – 1.5	Indicates consistency between observed data and Rasch model expectations	Bond, (2015); Boone, (2016)
	Item Fit (ZSTD)	−2.0 to +2.0	Detects statistically significant misfit	Linacre, (2012)
	Point-Measure Correlation (PTMEA CORR)	> 0.30	Indicates items align with the latent construct	
	Unidimensionality (PCA of residuals)	First residual contrast < 2.0 eigenvalue	Absence of substantial secondary dimensions	Linacre, (2012)
Reliability	Person Reliability	≥ 0.80	Consistency of person ability estimates	Bond & Fox, (2007)
	Item Reliability	≥ 0.80	Stability of item difficulty estimates	Boone, (2021)
	Person Separation Index	≥ 2.0	Ability to distinguish at least three ability strata	Wright & Masters, (1982)
	Item Separation Index	≥ 2.0	Ability to differentiate item difficulty levels	
Targeting & Precision	Person–Item Mean Difference	≤ ±1.0 logits	Indicates appropriate matching of item difficulty to ability	Bond & Fox, (2007)
Local Independence	Residual Correlations	< 0.30	Absence of local item dependence	Linacre, (2012)

Results

Based on the data collected from the e-learning system, there were 403 student test results. A unidimensionality assumption test was conducted as a prerequisite for data analysis. The results of the unidimensionality assumption test in this study are as follows:

Table 3. Results of unidimensionality test with principal component analysis (PCA)

Name	F1	F2	F3	F4	F5	H2	U2
Q1	0.29	-0.07	-0.05	0.23	-0.09	0.16	0.84
Q2	0.28	0.56	-0.11	-0.14	0.23	0.48	0.52
Q3	0.43	-0.13	-0.19	0.43	0.29	0.51	0.49
Q4	-0.41	0.00	0.32	-0.29	0.38	0.50	0.50
Q5	0.51	-0.09	0.08	0.07	0.26	0.35	0.65
Q6	0.27	-0.29	0.25	-0.28	-0.23	0.35	0.65
Q7	0.08	-0.36	0.32	-0.44	0.19	0.47	0.53
Q8	0.12	0.61	0.19	-0.20	-0.13	0.48	0.52
Q9	0.29	0.45	-0.32	-0.19	-0.31	0.53	0.47
Q10	0.51	-0.38	0.12	0.07	-0.35	0.55	0.45
Q11	-0.37	0.15	0.48	0.30	-0.35	0.60	0.40
Q12	-0.49	-0.03	-0.01	0.30	-0.40	0.49	0.51
Q13	-0.41	-0.16	-0.37	0.29	0.19	0.45	0.55
Q14	-0.19	0.09	0.06	0.54	0.47	0.56	0.44
Q15	-0.25	-0.24	-0.69	-0.10	-0.15	0.63	0.37

Value of the objective function = 0.0000

	F1	F2	F2	F3	F4	F5
Eigen value	1.86	1.40	1.40	1.32	1.27	1.24
Proportion Var	0.12	0.09	0.09	0.09	0.08	0.08
Proportion Explained	0.26	0.20	0.20	0.19	0.18	0.18

Based on Table 3 above, the results of Principal Component Analysis (PCA) of standardized residuals, these findings provide empirical evidence supporting the assumption of unidimensionality of the Pharmacy Mathematics assessment instrument. The eigenvalue for the first residual recorded is 1.86, which is below the critical threshold of 2.0, often used as an indication that there are no significant secondary dimensions in Rasch residual data (Smith et al., 2022). In addition, the proportion of variance explained by the first residual contrast is 12%, and the relatively small cumulative proportion of residual variance indicates that the residual pattern is more influenced by random error than by additional latent constructs. Therefore, the items in this instrument can be considered to consistently measure one main construct, namely the pharmaceutical mathematics abilities of students. These findings are in line with Rasch's methodological recommendations, which indicate that residual eigenvalues of less than 2.0 and low residual variance indicate a stable and construct-valid measurement structure (Linacre, 2012). Thus, these results reinforce the internal validity of the instrument and distinguish this study from classical test theory-based evaluation approaches, which generally do not provide direct evidence of residual-based unidimensionality.

The next assumption test was conducted by checking that each item was not correlated with each other. The results of the local independence assumption test based on the results of the Pearson correlation test with Jmetrik are as follows:

Table 4. Correlation matrix

	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Q8**	**Q9**	**Q10**	**Q11**	**Q12**	**Q13**	**Q14**	**Q15**
Q1	1.000	-0.1368	0.0514	-0.1338	-0.0554	-0.0984	-0.0297	-0.0101	0.0052	0.0571	-0.1202	-0.1179	-0.0952	-0.0475	-0.1292
Q2	-0.1368	1.000	0.0383	-0.0688	0.0135	-0.1246	-0.0625	0.0786	0.1368	-0.1047	-0.1626	-0.1503	-0.1441	-0.035	-0.1169
Q3	0.0514	0.0383	1.000	-0.2035	0.1032	0.016	-0.1387	-0.1138	-0.064	0.0462	-0.1766	-0.1559	-0.1288	0.0022	-0.0798
Q4	-0.1338	-0.0688	-0.2035	1.000	-0.1221	-0.0926	-0.0043	-0.0431	-0.1822	-0.1786	-0.0349	-0.0252	0.0095	-0.0533	-0.1081
Q5	-0.0554	0.0135	0.1032	-0.1221	1.000	-0.0573	0.0259	-0.0435	-0.0419	0.1298	-0.1635	-0.2287	-0.1906	-0.0376	-0.1397
Q6	-0.0984	-0.1246	0.016	-0.0926	-0.0573	1.000	0.0535	-0.0461	-0.0357	0.1081	-0.0938	-0.1192	-0.1037	-0.1398	-0.1267
Q7	-0.0297	-0.0625	-0.1387	-0.0043	0.0259	0.0535	1.000	-0.1929	-0.1021	-0.0068	-0.0865	-0.1354	-0.0592	-0.127	-0.101
Q8	-0.0101	0.0786	-0.1138	-0.0431	-0.0435	-0.0461	-0.1929	1.000	0.0698	-0.0863	0.0074	-0.1314	-0.0865	-0.1364	-0.1749
Q9	0.0052	0.1368	-0.064	-0.1822	-0.0419	-0.0357	-0.1021	0.0698	1.000	-0.0566	-0.1273	-0.1016	-0.1007	-0.0991	-0.0193
Q10	0.0571	-0.1047	0.0462	-0.1786	0.1298	0.1081	-0.0068	-0.0863	-0.0566	1.000	-0.0833	-0.1177	-0.1967	-0.1565	-0.0700
Q11	-0.1202	-0.1626	-0.1766	-0.0349	-0.1635	-0.0938	-0.0865	0.0074	-0.1273	-0.0833	1.000	0.1105	-0.1007	0.0011	-0.1554
Q12	-0.1179	-0.1503	-0.1559	-0.0252	-0.2287	-0.1192	-0.1354	-0.1314	-0.1016	-0.1177	0.1105	1.000	-0.0487	-0.0305	0.0023
Q13	-0.0952	-0.1441	-0.1288	0.0095	-0.1906	-0.1037	-0.0592	-0.0865	-0.1007	-0.1967	-0.1007	-0.0487	1.000	-0.0636	0.0903
Q14	-0.0475	-0.035	0.0022	-0.0533	-0.0376	-0.1398	-0.127	-0.1364	-0.0991	-0.1565	0.0011	-0.0305	-0.0636	1.000	-0.1131
Q15	-0.1292	-0.1169	-0.0798	-0.1081	-0.1397	-0.1267	-0.101	-0.1749	-0.0193	-0.0700	-0.1554	0.0023	0.0903	-0.1131	1.000

Based on Table 4 above, the analysis of local independence revealed through the residual correlation matrix between items, there is no indication of significant local dependence among the items in the Pharmacy Mathematics evaluation. All residual correlation values between item pairs are at a relatively low level, and none exceed the critical threshold of ±0.30, which is often proposed in the Rasch literature as a sign of local dependence. These results indicate that after controlling for the main latent construct effects with the Rasch model, students' responses to an item are not systematically influenced by their responses to other items, thus fulfilling the assumption of local independence. Fulfilling this assumption is crucial because violations of local independence can cause inflation of reliability and distortion in the estimation of item parameters and participant abilities (Zenisky et al., 2003). Therefore, these findings provide further evidence of the construct validity of the instrument and support the understanding that each item has a unique contribution in assessing students' pharmaceutical mathematics abilities, rather than merely reflecting similarities in content or question format.

Table 5. Rasch model analysis

Item	**Difficulty**	**Std. Error**	**WMS**	**Std.WMS**	**UMS**	**Std. UMS**
Q1	0.73	0.12	0.88	-2.57	0.78	-3.06
Q2	0.77	0.12	0.86	-2.96	0.76	-3.38
Q3	-1.50	0.13	0.96	-0.58	0.86	-1.27
Q4	-0.51	0.11	1.19	-4.31	1.24	3.66
Q5	0.61	0.11	0.85	-3.23	0.77	-3.57
Q6	0.74	0.12	0.86	-2.86	0.78	-3.13
Q7	-1.07	0.12	0.97	-0.67	0.96	-0.42
Q8	-1.03	0.12	0.92	-1.61	0.92	-0.96
Q9	1.75	0.14	0.89	-1.32	0.67	-2.52
Q10	1.53	0.13	0.85	-2.18	0.64	-3.29
Q11	-0.02	0.11	1.11	-2.71	1.15	2.69
Q12	0.30	0.11	1.18	-3.94	1.20	3.13
Q13	-0.63	0.11	1.13	-2.83	1.16	2.29
Q14	-1.23	0.12	1.06	-1.09	1.16	1.68
Q15	-0.44	0.11	1.29	-6.52	1.38	5.67

Table 6. Scale quality statistics

Statistic	Items	Persons
Observed Variance	0.9559	0.9541
Observed Std. Dev.	0.9777	0.9768
Mean Square Error	0.0140	0.3798
Root MSE	0.1183	0.6163
Adjusted Variance	0.9419	0.5742
Adjusted Std. Dev.	0.9705	0.7578
Separation Index	82.074	12.296
Number of Strata	112.765	19.728
Reliability	0.9854	0.6019

Analysis of item characteristics using the Rasch Model shows that, overall, all fifteen items of the Pharmacy Mathematics assessment instrument meet the basic assumptions in measurement, which include unidimensionality and local independence. These results indicate that after controlling for students' latent abilities through the model, responses to one item are not influenced by responses to other items. Compliance with the assumption of local independence is crucial, because violation of this assumption can result in increased reliability and bias in parameter estimation, thereby making the measurement results less valid in terms of construct.

In terms of item difficulty, the difficulty values, indicating sufficient variation in difficulty to measure student abilities from low to high levels. Items with negative difficulty levels, such as items 3, 7, 8, and 14, serve as easy items that are effective in measuring basic abilities and preventing the floor effect, especially among students with low initial abilities. Conversely, items with high difficulty levels, for example, items 9 and 10, with values > 1.5, play an important role in distinguishing students with higher mathematical abilities. This balanced difficulty distribution reflects the quality of the instrument in providing measurement information across the entire range of abilities, as recommended in the Rasch literature for educational assessment (Wright & Masters, 1982).

Based on the model fit results, most items show Weighted Mean Square (WMS)/INFIT which reflects the extent to which responses correspond to items related to the respondent's abilities, thereby identifying inconsistent response patterns on items of moderate difficulty. Meanwhile Unweighted Mean Square (UMS)/OUTFIT mean square (MNSQ) which describes the consistency of responses to questions that are far from the respondent's level of expertise, making it more sensitive to extreme or outlier responses, values within the acceptable range of 0.5–1.5, indicating that the students' response patterns are generally consistent with the Rasch model expectations. Several items, such as items 1, 2, 5, and 6, show slight overfit with MNSQ < 1, indicating highly predictable responses and generally associated with a relatively simple item structure; this condition is not considered substantively problematic. Conversely, several other items, such as items 4, 12, and 15, show slight underfitting with MNSQ values approaching the upper limit of the criteria, which may reflect variation in solution strategies, context ambiguity, or differences in student interpretation. Nevertheless, these items are still worth retaining with recommendations for editorial revision or further content review. Overall, the combination of evidence of local independence, adequate difficulty variation, and good model fit confirms that this instrument is valid and reliable for measuring students' mathematical abilities in the context of Pharmaceutical Mathematics (Embretson & Reise, 2000).

Based on the output of the Rasch model Item Characteristic Curve (ICC) analysis, it can show items with the highest quality and items with the lowest quality.

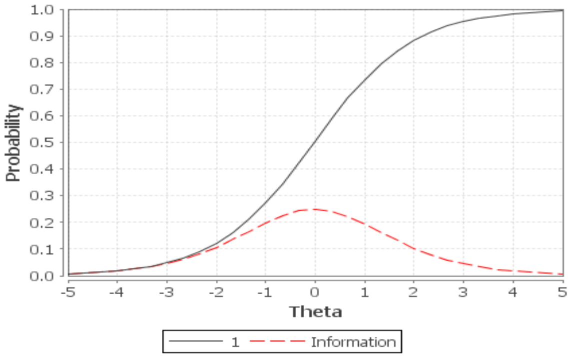


Figure 2. Item characteristic curve (ICC) question 11

Based on the results of the local independence test and item difficulty mapping on the logit scale, Rasch analysis shows that item 11 is the item with the highest psychometric quality in this instrument. This item is very close to the midpoint of the ability scale (around 0 logit), which coincides with the average ability of respondents, thus providing the most optimal measurement information. In addition, the residual correlation of item 11 with other items is well below the critical limit of ±0.30, indicating no local dependence and that responses to this item are entirely determined by the latent ability being measured. In the Rasch framework, items located around the ability mean and meeting the assumption of local independence are considered core items because they greatly contribute to the stability of ability estimates and overall measurement precision.

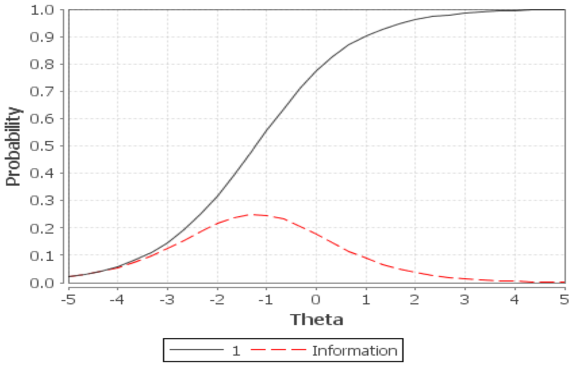


Figure 3. Item characteristic curve (ICC) question 14

Based on Figure 3 for item 14 can be identified as the item with the lowest relative quality, although it is still within the model's acceptability limits. This item is located on a negative logit that is quite far from the center of the ability distribution, indicating that its level of difficulty is very low and only provides meaningful information for respondents with basic abilities.

Although the residual correlation between items remains below the ±0.30 threshold, which means that the assumption of local independence is still fulfilled, the external contribution of this item to the overall measurement precision is relatively limited. From a modern measurement perspective, items that are too easy tend to be less sensitive in distinguishing abilities at the intermediate to high levels and have the potential to reduce the efficiency of the instrument if not balanced with items of moderate and high difficulty. Therefore, although item 14 is still useful as a screening item for basic abilities, editorial revision or adjustment of the difficulty level is recommended to improve its contribution to the overall quality of measurement.

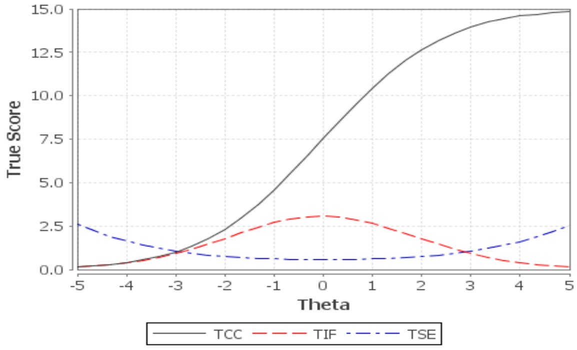


Figure 4. Person plot instruments

The Person Plot graph in IRT analysis provides a comprehensive view of how measurement tools operate in assessing respondents' latent abilities (theta). This graph usually consists of three main elements, namely the Test Characteristic Curve (TCC), Test Information Function (TIF), and Test Standard Error (TSE). These three elements are interrelated and form the fundamental basis for assessing the suitability of measurement tools with the distribution of respondents' abilities. The TCC curve illustrates the relationship between the latent ability of respondents (θ) and the expected true score. In the graph presented, the TCC shows a sigmoid shape that increases consistently from left to right. This pattern is consistent with the basic assumptions of the IRT and Rasch models, which state that the higher the latent ability of respondents, the greater the likelihood that they will answer items correctly, resulting in a regular increase in true scores.

The TIF curve illustrates the amount of information generated by the test at each level of latent ability. In the context of IRT, this information has an inverse relationship with measurement error, meaning that the higher the TIF value, the more accurate the estimation of the respondent's ability. In the Person Plot diagram presented, TIF reaches its highest point between $\theta = 0$ and $\theta = +1$. This range is in line with the center of the respondent's ability distribution, indicating that the measuring instrument is designed and functions most optimally to evaluate abilities that are at an average level to slightly above. From a psychometric point of view, this condition is highly desirable in educational assessment, as most respondents usually fall within that ability range.

The TSE curve is the inverse of the TIF, because mathematically, TSE is a function of inverse test information. In the graph, it can be seen that TSE reaches its lowest value in the ability range that coincides with the maximum point of TIF, which is located around $\theta = 0$ to θ

= +1. This indicates that measurement errors are minimal at average abilities, so estimates of respondents' ability levels in this range are the most accurate.

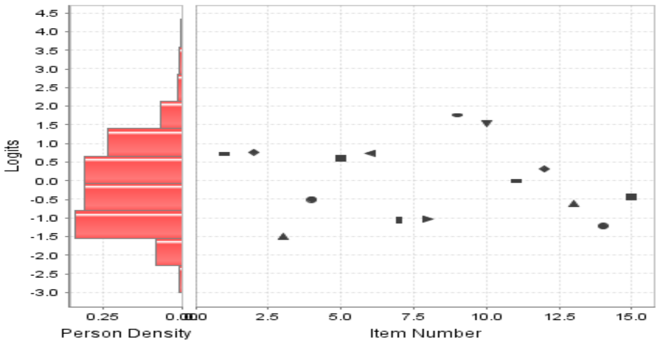


Figure 5. Item map

The Item Map or Wright Map presents simultaneous information about the distribution of participants' abilities (individual density) and the level of difficulty of questions on a similar logit scale. Thus, question analysis is not only based on difficulty parameters, but also on the relevance of their position to the distribution of participants' abilities. Based on this map, the fifteen questions show a fairly even distribution of difficulty from low to high logits, although there appears to be a concentration of participant abilities in the middle to lower range of the logit scale.

Discussion

This study aims to assess the psychometric quality of mathematics assessment tools using the Item Response Theory approach, specifically the Rasch Model, to ensure that these tools can measure students' abilities in a valid and objective manner. In general, the results of the study show that this objective has been achieved, as reflected in the model fit for most items, the stability of the difficulty hierarchy, and the logical structure of the measurement information distribution across the ability range. In relation to the initial objective of the study, the results of the analysis show that the items function in accordance with the intended mathematical ability construct. The variation in item difficulty levels allows this tool to differentiate students with abilities ranging from low to high, thus supporting the basic assumption of the Rasch Model regarding the mapping of respondent abilities and item difficulty on the same logit scale. These findings reinforce the argument that the Rasch approach is more effective than classical test theory in providing sample independent measurement interpretations (Muhtarom, 2024).

Scientific interpretation of the findings shows that the dominance of items with good fit reflects the correspondence between students' empirical responses and the theoretical expectations of the model. Slight overfit on some items can be understood as a signal that these items are highly structured and procedural, making students' responses fairly easy to predict. Conversely, slight underfitting on some other items indicates variation in cognitive strategies or ways of understanding the context of the question. In the modern psychometric perspective, this condition does not always have a negative meaning, but can reflect complexity in mathematical thinking processes, especially on questions that require conceptual understanding and contextual application (Mcneish & Wolf, 2020).

The results of the person fit and test information function analyses indicate that this instrument provides the most optimal measurement information in the range of average to slightly above average abilities. This pattern is consistent with the findings of many previous Rasch studies that report that the highest level of measurement precision is usually found in the range of abilities where the majority of respondents are located (Boone, 2021). Thus, the findings in this study support previous research and do not contradict existing theory. However, the limited information at the extremes of ability also emphasizes the need to develop additional items that specifically target very low and very high abilities.

This study reinforces and expands the psychometric evidence reported in recent Rasch studies by presenting a more comprehensive and diagnostic analysis that goes beyond simply reporting the level of difficulty and model fit predominantly found in previous literature. Based on the results of Wai et al (2024), the use of the Rasch model in instrument development is not only to assess the validity and reliability of reading instruments but also to combine findings with graphical representations of person item maps to visually demonstrate the suitability between items and respondents' abilities, while (Hermansyah et al., 2024) applied the Rasch model in their research by including DIF analysis to identify potential biases based on demographics in school welfare measurement tools. Unlike the two studies mentioned above, this study not only reports fit statistics and item difficulty levels, but also presents a person-item map that illustrates the distribution of items across the abilities of pharmacy students, thereby empirically demonstrating whether the instrument covers the relevant range of abilities and detects gaps or item redundancy. Furthermore, although some studies focus on statistical fit, this approach is designed to anticipate further Differential Item Functioning (DIF) as part of a comprehensive evaluation strategy, thereby providing stronger evidence of fairness and equity of measurement among respondent subgroups, for example, based on gender or study program, which directly addresses the modern research need for evaluation methods that are sensitive to group differences. Thus, the integration of person-item mapping and bias detection within the Rasch framework enhances the empirical contribution of this study compared to other research, while providing a stronger basis for the use of valid, reliable, and fair Pharmacy Mathematics assessment instruments across a wide range of abilities.

The results of the Rasch model analysis in this study indicate that the Pharmacy Mathematics assessment instrument meets the main criteria of modern measurement, including adequate model fit, unidimensionality, and local independence. First, the range of item difficulty, which ranged from −1.50 to 1.75 logits, indicates that although the instrument covers low to high abilities, the targeting of items for students with very high abilities is still relatively limited. This is reflected in the concentration of several difficult items in the adjacent logit range, which has the potential to reduce measurement precision at the upper tail of the ability distribution.

In addition, another limitation relates to the scope of the Rasch analysis conducted. This study did not formally implement item bias analysis Differential Item Functioning (DIF), even though the sample composition showed differences in demographic characteristics (e.g., gender and type of study program). Thus, the potential for measurement bias between subgroups could not be completely eliminated. Furthermore, the use of data from a single institutional context limits the generalization of findings to a broader population of pharmacy students. Therefore, the results of this study should be interpreted as preliminary evidence of the quality of the instrument, which still requires replication in a more diverse sample and further Rasch analysis, such as DIF and cross context person item targeting, to strengthen external validity. The open disclosure of these limitations is intended to increase transparency and provide clear direction for future research in the development of Rasch-based Pharmacy Mathematics assessment instruments.

Conclusion

The Pharmacy Mathematics assessment instrument for students shows strong and coherent measurement quality in measuring one latent construct, namely pharmacy mathematics ability. All items meet the model fit criteria with INFIT and OUTFIT Mean Square values within acceptable limits, supported by evidence of unidimensionality and local independence, which confirms the construct validity of the instrument. The distribution of item difficulty levels shows adequate variation to measure student abilities from low to high levels, where easy items serve to prevent floor effects and difficult items serve to distinguish students with high abilities. However, the variation in informational contribution between items shows that some items still provide limited measurement precision. Overall, these findings confirm that the instrument has adequate validity and reliability as a tool for measuring students' Pharmacy Mathematics abilities, while also providing a clear empirical basis for the development of more precise and equitable instruments in future research.

Acknowledgment

The authors would like to thank all parties who have assisted in the research process leading up to the publication of this article, particularly the IKIFA College of Health Sciences, the students who volunteered to be samples in this study, and the Research Team.

Declarations

Conflicts of Interest	:	The authors declare that there are no conflicts of interest related to the publication of this manuscript. In addition, all ethical aspects, including but not limited to plagiarism, professional misconduct, falsification and/or manipulation of data, duplicate publication and/or submission, and redundancy, have been thoroughly addressed and resolved by the authors.
Generative AI Statement	:	Generative AI tools, such as DeepL and Scopus AI, were employed solely for language editing and minor phrasing enhancements. All conceptualization, analysis, and scholarly content were independently developed and verified by the authors.

Funding Statement	:	This scientific research received financial support through internal research funding allocated from the IKIFA College of Health Sciences budget, as stipulated in the Decree of the Head of STIKes IKIFA Number 017-02/SK/STI-IKIFA/XII/2024.
Author Contributions	:	**Alifa Sabrina:** Conceptualization, Writing-Original Draft, Methodology, Formal Analysis; **Fitri Savitri:** Writing-Review & Editing, Visualization, Project Administration; **Farida Tuahuns:** Software, Validation, Data Curation.

References

Bond, T. G. (2015). *This is the preface from the first edition of Bond & Fox (2001). It outlines the principles and practices which drove us to write the book. The third edition (2015).* www.routledge.com/9780415833424.2001.

Bond, T. G., & Fox, C. M. (2007). *Applying the rasch model : fundamental measurement in the human sciences second edition university of toledo*.

Boone, W. J. (2016). *Rasch analysis for instrument development : why , when , and how ?* https://doi.org/10.1187/cbe.16-04-0148

Boone, W. J. (2021). Rasch measurement: applications in health professions education. *Health Professions Education*, *7*(3), 275–284.

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American Journal Of Medicine*, *119*(2), 521–529. https://doi.org/10.1016/j.amjmed.2005.10.036

Crocker, & Algina. (2008). *Introduction to classical and modern test theory. Cengage Learning.* Cengage Learning.

De Champlain, A. F. (2010). A Primer on classical test theory and item response theory for assessments in medical education. *Medical Education*, *44*(1), 109–117. https://doi.org/10.1111/j.1365-2923.2009.03425.x

Derendorf, H., & Schmidt, S. (2020). *Clinical pharmacokinetics and pharmacodynamics rowland and tozer's*.

Dwijayanti, I., Utami, R. E., & Nugroho, A. A. (2021). Peningkatan kemampuan tpack calon guru: analisis kemungkinan kesalahan pemahaman konsep aljabar serta strategi pencegahannya [*Improving prospective teachers' tpack competence: analysis of possible misunderstandings of algebraic concepts and prevention strategies*]. *Aksioma*, *10*(4), 2283-2290. https://doi.org/10.24127/ajpm.v10i4.4119

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates, Inc., Publishers.

Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: principles and applications. In *Applied Psychological Measurement* (Vol. 9, Issue 3). Springer Science. https://doi.org/10.1177/014662168500900315

Haute, A. de S. (2021). *Medicinal product dose calculation*. 1–3.

Hermansyah, E., Rizkyanti, C. A., & Hakim, L. N. (2024). Rasch calibration and differential item functioning (DIF) analysis of the school well-being scale for students Eka. *Jurnal Kependidikan*, *10*(1), 1–11.

Linacre, M. (2012). *A user's guidelines winstep*. https://www.winsteps.com/winman/index.html

Mcneish, D., & Wolf, M. G. (2020). *Thinking twice about sum scores*. *2020*(April), 2287–2305.

Muhtarom, M. (2024). Developing an instruments to measure prospective teacher beliefs about mathematical problem-solving using the Rasch model. *Jurnal Elemen*, *10*(2), 274–288.

https://doi.org/10.29408/jel.v10i2.25040

Mulac, A., Hagesaether, E., & Granas, A. G. (2022). Medication dose calculation errors and other numeracy mishaps in hospitals: Analysis of the nature and enablers of incident reports. *Journal of Advanced Nursing*, *78*(1), 224–238. https://doi.org/10.1111/jan.15072

Prasetya, W. A. (2023). Item quality analysis using the rasch model to measure …. *JPEP*. https://scholarhub.uny.ac.id/jpep/vol27/iss1/7/

Salafiya, T. T., Kamila, F. I., Kimnu, K. P., & Risqiyani, L. (2025). *Kesulitan mahasiswa pendidikan matematika Universitas Negeri Semarang dalam memahami konsep kalkulus integral [Difficulties faced by mathematics education students at Semarang State University in understanding the concept of integral calculus]*. Jurnal Angka, *2*(1), 93–108.

Smith, S. K., Benbenek, M. M., Bakker, C. J., & Bockwoldt, D. (2022). Scoping review: Diagnostic reasoning as a component of clinical reasoning in the U . S . *Primary Care Nurse Practitioner Education, June*, 3869–3896. https://doi.org/10.1111/jan.15414

Sugiyono. (2014). *Metode penelitian bisnis: Pendekatan kuantitatif, kualitatif, kombinasi, dan R&D [Business research methods: Quantitative, qualitative, mixed-methods, and R&D approaches].* Alfabeta.

Wagner, J. G. (2020). Pharmacokinetics for the pharmaceutical scientist. In *pharmacokinetics for the pharmaceutical scientist*. Springer. https://doi.org/10.1201/9780203743652

Wai, L., Yim, K., Lye, C. Y., & Koh, P. W. (2024). *A psychometric evaluation of an item bank for an English reading comprehension tool using Rasch analysis*. *10*(1), 18–34.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. MESA Press.

Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2003). Effects of local item dependence on the validity of irt item, test, and ability statistics. University of Massachusetts at Amherst. *Medical College Admission Test*.

Appendix

No.	Indicators	Question Items (in Bahasa)
1	Menerapkan fungsi eksponensial dan logaritma dalam farmakokinetika	1. Skala pH didefinisikan sebagai pH = -log[H^+]. Jika konsentrasi ion H^+ dalam suatu larutan 3,162 x 10^{-5} M, pH larutan tersebut adalah ….. (Note: log 3,162 = 0,5) A. 3,5 B. 4,0 C. 4,5 D. 5,0 E. 5,5
		2. Sebuah obat mengalami dekomposisi mengikuti kinetika orde pertama. Jika konsentrasi awal adalah 100 mg/mL dan konstanta laju dekomposisi adalah 0,05 per hari. Konsentrasi obat setelah 10 hari adalah …. (Note: $e^{-0,5}$= 0,6065) A. 50,65 mg/mL B. 56,05 mg/mL C. 60,65 mg/mL D. 65,05 mg/mL E. 66,05 mg/mL
		3. Seorang petugas farmasi ingin mengetahui waktu paruh (t½) suatu obat eliminasi orde pertama. Jika konstanta laju eliminasi (k) adalah 0,1155 per jam, waktu paruh obat tersebut adalah …. (Rumus: t½ = $\frac{\ln 2}{k}$; In 2 = 0,6930) A. 2 jam B. 4 jam C. 6 jam D. 8 jam E. 10 jam
		4. Konsentrasi obat berubah menurut $C = 40(1 - e^{-0.3t})$. Setelah 4 jam, konsentrasi obat adalah …. (Note: $e^{-1,2} = 0{,}30$) A. 25 B. 28 C. 31 D. 34 E. 37
2	Menggunakan konsep aljabar	5. Rumus penghitungan laju eliminasi obat: $R = kC$, di mana $k = 0.15$dan $C = 40 - 5t$. Besar nilai R saat $t = 4$ adalah …. A. 2.0 B. 3.0 C. 4.5 D. 5.0 E. 6.0

No.	Indicators	Question Items (in Bahasa)
	dalam penyelesaian persamaan farmasi	6. Persamaan kadar obat terhadap waktu adalah $C = 50 - 5t$. Setelah …. jam kadar obat menjadi 25 mg/L. A. 3 jam B. 4 jam C. 5 jam D. 6 jam E. 7 jam
		7. Persamaan kadar obat dinyatakan $C = \frac{200}{t+5}$. Jika kadar obat yang diinginkan adalah 20 mg/L, pemberian dosis obat tersebut tercapai, pada saat t = …. jam. A. 5 jam B. 10 jam C. 15 jam D. 20 jam E. 25 jam
3	Menerapkan perbandingan dan proporsi dalam perhitungan dosis	8. Seorang pasien memerlukan obat dengan dosis $D = 2x + 5$mg, di mana $x$adalah berat badan pasien dalam kg dibagi 10. Jika berat pasien 70 kg, dosis obat yang diberikan adalah …. A. 13 mg B. 15 mg C. 17 mg D. 19 mg E. 21 mg
		9. Dosis obat dihitung dengan rumus $D = \frac{mg \times BB}{100}$. Jika dosis standar 150 mg untuk berat badan 50 kg, dosis yang diberikan untuk berat badan 80 kg adalah …. A. 200 mg B. 220 mg C. 240 mg D. 250 mg E. 300 mg
4	Menggunakan konsep larutan dan pengenceran	10. Jika volume larutan akhir V_2diperoleh dari pengenceran $C_1V_1 = C_2V_2$, maka $V_2 = \frac{C_1V_1}{C_2}$. Besar nilai V_2jika $C_1 = 10\%$, $V_1 = 25\ mL$, dan $C_2 = 2\%$ adalah …. A. 50 mL B. 100 mL C. 125 mL D. 200 mL E. 250 mL
		11. Agar dapat mencampur dua larutan 20% dan 10% agar diperoleh larutan 15% sebanyak 100 mL, volume larutan 20% yang dibutuhkan adalah …. A. 25 mL B. 40 mL C. 45 mL D. 50 mL E. 60 mL

No.	Indicators	Question Items (in Bahasa)
		12. Agar dapat mengurangi konsentrasi larutan dari 12% menjadi 8%, banyak air (tanpa zat aktif) harus ditambahkan ke 200 mL larutan awal adalah …. A. 50 mL B. 75 mL C. 100 mL D. 200 mL E. 300 mL
5	Menerapkan konsep matematika dasar dalam konteks farmasi dan fisika	13. Hubungan antara volume (V) dan tekanan (P) gas anestesi memenuhi $PV = k$. Jika $P = 2\ atm$ dan $V = 5\ L$, agar besar tekanan tetap sama, saat volume menjadi 8 L adalah …. A. 0.5 atm B. 1.0 atm C. 1.25 atm D. 2.0 atm E. 3.0 atm
		14. Jika kadar obat meningkat linier dengan waktu $C = 5t + 10$, kadar setelah 6 jam adalah …. A. 20 mg/L B. 30 mg/L C. 35 mg/L D. 40 mg/L E. 45 mg/L
		15. Dalam suatu rumus aljabar farmasi, $C = \frac{a+b}{2}$. Jika kadar dua senyawa berturut-turut 40 mg/L dan 60 mg/L, maka kadar rata-rata adalah …. A. 40 mg/L B. 45 mg/L C. 50 mg/L D. 55 mg/L E. 60 mg/L