Al-Ishlah: Jurnal Pendidikan
Vol.17, 3 (September, 2025), pp. 4838-4847
ISSN: 2087-9490 EISSN: 2597-940X, DOI: 10.35445/alishlah.v17i3.6411

Exploring Number Concept Mastery: A RASCH Analysis of Flipped Classroom Learning for Future Elementary Teachers

Irni Rachmawati Putri[1], Rahayu Condro Murti[2], Fery Muhamad Firdaus[3]

[1] Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; irnirachmawati.2023@student.uny.ac.id
[2] Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; rahayu_cm@uny.ac.id
[3] Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; fery.firdaus@uny.ac.id

ARTICLE INFO

Keywords:

RASCH modeling;
flipped classroom;
number concept

Article history:

Received 2024-11-28
Revised 2025-02-07
Accepted 2025-09-29

ABSTRACT

Mastering the concept of numbers is a critical competency for prospective elementary school teachers, as it forms the foundation of early mathematics instruction. This study explores the effectiveness of the flipped classroom model in enhancing number concept understanding, supported by RASCH analysis to measure and interpret student performance. A quantitative descriptive design was employed involving 39 first-semester students from Class D of the Elementary Teacher Education (PGSD) Program at Yogyakarta State University (Class of 2024). Data were collected using a multiple-choice test on number concepts, covering topics such as types of numbers, arithmetic operations, and number properties. The results were analyzed using the RASCH Model through WINSTEPS software, focusing on Wright Map, item fit, student ability, scalogram analysis, and test reliability. The RASCH analysis revealed that the flipped classroom model supported diverse levels of student understanding. The Wright Map indicated varied ability levels across students, while item fit statistics confirmed that most items functioned as expected. Scalogram analysis identified response patterns indicating careless mistakes, lucky guessing, and potential cheating. The test demonstrated high reliability in distinguishing between ability levels. Findings suggest that the flipped classroom model contributes positively to the conceptual understanding of number concepts, though learning gaps remain. The RASCH Model provided in-depth diagnostic insights, allowing for more targeted instructional improvements. The integration of flipped learning with RASCH-based assessment offers a robust framework for improving and evaluating mathematics instruction among prospective teachers.

Corresponding Author:
Irni Rachmawati Putri
Universitas Negeri Yogyakarta, Yogyakarta, Indonesia;;irnirachmawati.2023@student.uny.ac.id

1. INTRODUCTION

Mathematics learning is a universal field of science and has characteristics essential for the development of education at various levels, from elementary school to college (Fauzi et al., 2020). At the elementary school level, mathematics learning not only aims to introduce fundamental concepts but also has a central role in shaping students' mindsets so that they can think critically, creatively, logically, and systematically (Rahayu & Kusuma, 2019). This is very important, considering that mathematics in elementary school is the foundation for understanding more complex concepts at the next level of education (Aledya, 2020). Mathematics learning is also considered one of the fields of study with a high level of difficulty, mainly because the concepts taught are often abstract and require a contextual approach to be relevant to students' daily lives (Fauzi, 2022). Therefore, educators need to package mathematics learning with a holistic and integrative approach so that students can understand the concepts taught better (Rusilowati, 2018).

Mathematics learning materials at the elementary school level include various fundamental concepts, such as addition, multiplication, division, and measurement, which, as a whole, are often referred to as learning the concept of numbers (Sutisnawati, 2017). This number concept is relevant at the elementary school level and becomes the foundation for understanding further mathematics materials such as arithmetic, algebra, and statistics (Rahayu, 2021). Therefore, the concept of numbers is the main focus of mathematics learning in elementary school, and it has a crucial role in building students' understanding of mathematics in general. This learning must be carried out using the proper methods and strategies so that students can understand the concept deeply.

In the context of elementary school teacher education, student teachers need to have a strong understanding of the concept of numbers because this concept will be an integral part of the mathematics learning process that they teach in the future. One of the learning models currently considered adequate for teaching the concept of numbers is the flipped classroom model. This model is part of the Blended Learning approach that combines online and offline learning, allowing students to better understand the material through contextual and personal learning (Sariningsih & Purwasih, 2017). The flipped classroom model allows students to study primary material independently first before discussions or learning activities in class take place. Thus, learning carried out in the classroom can be focused on more interactive and in-depth activities, which help students construct their understanding more meaningfully (Sutisnawati, 2017).

Along with the development of learning methods, measuring students' conceptual understanding has become a very important aspect of the learning process. Measurement in the context of education is a process carried out by teachers to obtain information about how far students understand the material being taught (Yuwono et al., 2018). In this context, measurement can be carried out through various instruments, one of which is a test specifically designed to assess students' understanding of certain concepts (Tarigan et al., 2022). This measurement instrument must be designed so that it is relevant to the learning objectives and conditions of the students (Dwinata, 2019). These tests are not only used to measure students' knowledge but also to assess their skills and attitudes towards learning mathematics.

One of the methods used to conduct an in-depth analysis of test results is the RASCH Model. This model provides a more complex analytical approach to the measurement process by allowing the identification of inappropriate or uncommon test items (misfits and outliers) and providing replicable measurement results (Suryani, 2018). In addition, the RASCH Model also provides a clear picture of the relationship between student abilities and the difficulty of the questions faced (Tyas et al., 2020). In other words, this model provides more detailed insight into the level of student understanding of the concepts taught and helps teachers evaluate the effectiveness of the instruments used in the measurement process.

However, using conventional learning models often poses challenges for teachers, especially in accurately measuring student understanding. In many cases, students tend to use guessing or cheating techniques when working on problems so that the assessment results do not reflect their proper understanding (Putra et al., 2018). Research conducted by Nuryadin (2022) shows that using a blended learning model, which combines online and offline approaches, can help overcome this problem. The study proved that applying a blended learning model in geometry learning in elementary schools improved students' understanding of the mathematical concepts taught. The results of this study also show that this approach can be applied effectively in mathematics learning at the tertiary level, especially for prospective elementary school teacher students.

Therefore, prospective elementary school teachers need to master basic mathematical concepts, especially the concept of numbers, which will be the foundation for mathematics learning in elementary school. By implementing the flipped classroom model and using the RASCH Model analysis, prospective elementary school teachers can develop more effective and meaningful learning strategies. The flipped classroom model allows students to master concepts independently, while the RASCH Model provides an accurate evaluation tool to measure student understanding. RASCH modeling is able to provide in-depth analysis by providing clear information and descriptions of students' conceptual understanding through analysis of the instruments used to measure students' abilities. Based on this background, researchers are interested in analyzing the ability to understand the concept of the number of prospective elementary school teachers who apply learning with the flipped classroom model through RASCH modeling analysis.

2. METHODS

This research is a quantitative descriptive study that focuses on the analysis of test instruments using WINSTEPS software. The study subjects were prospective elementary school teachers in semester 1 from class D, PGSD Study Program, Yogyakarta State University, class of 2024. The students who were the subjects of the study were between 17 and 18 years old. The research employed a purposive sampling technique, selecting 39 students based on their enrollment in the introductory number concept course. The justification for this sample size is based on guidelines for Rasch modeling, which suggest a minimum of 30 respondents for stable parameter estimation (Linacre, 1994). Further justification of the sample adequacy is needed to ensure robust results. The test instrument used in this study was a multiple-choice test designed to assess students' understanding of number concepts. The questions consisted of eight items covering the material of the origin of numbers, types of numbers, number arithmetic operations, and the properties of number arithmetic operations. The difficulty level of the questions was categorized into three levels: easy, medium, and challenging. Before administering the test, the instrument underwent content validation by a panel of experts in mathematics education. The validation process involved expert review, where three mathematics educators evaluated the test items for content relevance, clarity, and alignment with learning objectives; pilot testing, where the test was administered to a small group of students who were not part of the main study to identify potential ambiguities and improve item formulation; and reliability analysis, where the internal consistency of the test was measured using Cronbach's Alpha, and item reliability was assessed using WINSTEPS software.

Data processing was conducted by scoring the students' answer sheets using the Guttman scale, where correct answers received a score of 1 and incorrect answers received a score of 0 (Sugiyono, 2013). The scored data were then entered into Microsoft Excel and exported in *prn file format, which could be accessed via Notepad. Subsequently, the data were imported into the WINSTEPS software for Rasch modeling analysis. The Rasch analysis included Wright Map Analysis (Person-Item Map), which provided insights into the distribution of student abilities and question difficulty levels on the same scale; Item-Measure Analysis, which evaluated the fit of each question item in measuring student ability and identified misfit items; and Scalogram Analysis, which classified students into three types: careless students, lucky students (lucky guessers), and students who engaged in answer sharing. WINSTEPS software was also used to measure test reliability and validity through item reliability, where the separation index and reliability coefficient were calculated to ensure the test could distinguish between different ability levels; person reliability, which assessed the consistency of student responses across different test items; and item fit statistics, where infit and outfit mean square statistics were analyzed to identify items that did not conform to the expected Rasch model. By incorporating these methodological enhancements, the study ensures a more rigorous and comprehensive analysis of the test instrument's effectiveness in measuring students' understanding of number concepts.

3. FINDINGS AND DISCUSSION

3.1 WRIGHT Map Analysis

WRIGHT map analysis, also known as person item map, evaluates the distribution of students' ability levels and the difficulty level of the questions. This study analyzes the ability of 39 students in class D studying in the first semester of the PGSD study program at Yogyakarta State University to understand basic mathematical concepts with data entered the RASCH modeling system. Through the RASCH distribution map, the level of student ability is depicted on the left side, while the difficulty level of the questions is displayed on the right side. The distribution of student ability on the right side of the map includes an analysis of the level of expertise of 39 students in answering all the test questions given. This distribution map shows the results of student work related to geometry material with a flipped classroom learning approach, as depicted in Figure 1 below.

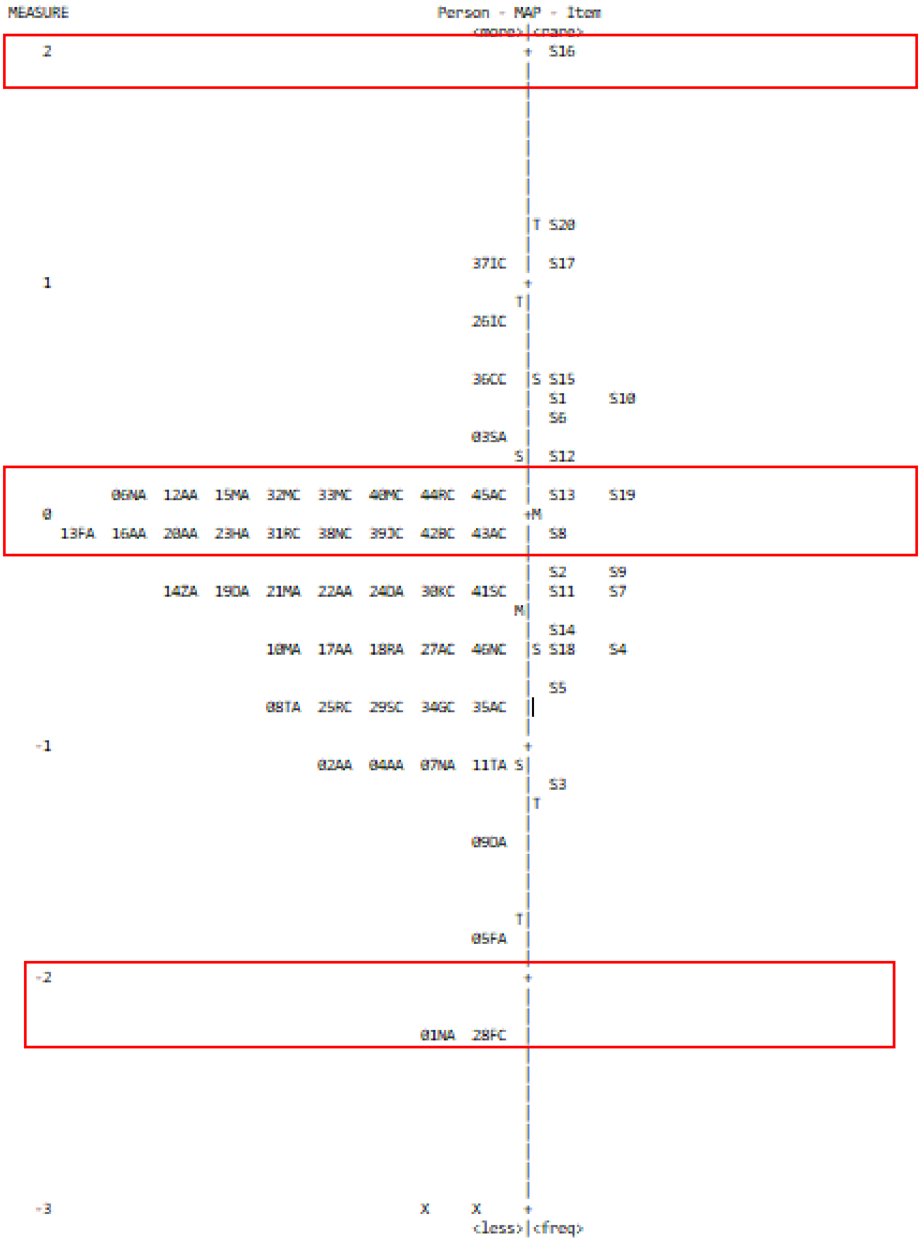

Figure 1. WRIGHT Map

The intervals at Mean, -1 SD, and 1 SD represent the level of variation in question difficulty. Based on the analysis of the distribution map, the question with the highest level of difficulty is S16 (question number 16). In contrast, the question with the lowest level of difficulty is S3 (question number 3). The distribution map in Figure 1 shows that the distribution of students' abilities in question difficulty has a broader range than the distribution of question difficulty levels. This illustrates that the abilities of 39 class D PGSD students at Yogyakarta State University in the material on basic mathematical concepts are very diverse. Based on the map, students with the highest classification have an ability value of +2.0 logit. Students who fall into this category, such as Student 37IC, are outside the +1SD deviation, indicating they have very high abilities. The average person logit value of +1.00 logit with an average item logit of 0.00 logit indicates that most students understand mathematical concepts. However, there is still a need to improve learning methods. This aims to reduce significant gaps in ability among students, considering that each individual has different needs and focuses. Thus, although the average student's ability is in line with easy questions, their ability level still needs to fully meet the optimal standard for questions with a higher difficulty level.

The analysis results show that the intervals at Mean, -1 SD, and +1 SD represent the level of variation in question difficulty that correlates with the distribution of student abilities. The distribution map shows that the question with the highest difficulty level is S16 (question number 16), while the question with the lowest difficulty level is S3 (question number 3). The distribution of student abilities is more varied compared to the level of question difficulty, reflecting significant differences in understanding basic mathematical concepts among class D PGSD students at Yogyakarta State University. This is supported by Wright's (1997) findings, which state that the level of heterogeneity in a study group can cause significant gaps in learning outcomes, especially in conceptual material.

The highest ability of students, with an ability value of +2.0 logit, is outside the +1 SD deviation, as achieved by Student 37IC. This indicates that some students have very good abilities in understanding basic mathematical concepts. However, the average logit person value of +1.00 logit, which is lower than the highest ability value, indicates that most students are at a pretty good level of ability but still need to be optimal in solving problems with a higher difficulty level. According to previous research by Widhiarso (2015), these results reflect the need for a variety of learning approaches that can increase students' absorption of more complex concepts so that they can overcome problems with a high level of difficulty.

In addition, significant ability gaps among students indicate the need to evaluate the learning methods used. Each individual has different learning needs and focuses, so a more adaptive approach, such as blended learning or project-based learning, can help reduce these gaps. According to Syadiah and Hamdu (2020), learning tailored to individual needs can improve understanding skills and motivate students to achieve higher learning standards. Thus, efforts to improve teaching strategies are essential to ensure that the average ability of students is in line with the difficulty level of questions that are by optimal standards.

3.2 Item Measure Analysis

Item analysis, often called item measures in RASCH modeling, can be evaluated through data processing and interpretation of the item measure output table. This analysis process aims to identify the contribution of each test instrument or question in measuring students' abilities, especially in solving problems in geometry concepts (Dwinata, 2019). This modeling produces a table that provides an overview of each question's difficulty level. This information includes various categories of item difficulty levels (Yulianto & Widodo, 2020). A critical component of this analysis is the table on the JMLE Measure and items. Based on Figure 1, it is known that the question with the highest level of difficulty is question number 16 (S16), with a logit value of 4.78. Conversely, the question with the lowest difficulty level is question number 3 (S3), which has a logit value of -1.20.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE 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
16	0	48	4.78	1.80	MAXIMUM MEASURE				.00	.00	100.0	100.0	S16
20	8	48	1.27	.40	1.12	.56	1.26	.81	.08	.21	82.6	82.6	S20
17	9	48	1.12	.38	.93	-.27	.92	-.15	.28	.22	80.4	80.4	S17
15	13	48	.61	.34	1.03	.22	.96	-.10	.25	.26	71.7	72.5	S15
1	14	48	.50	.33	1.07	.56	1.00	.08	.22	.27	63.0	70.6	S1
10	14	48	.50	.33	1.02	.20	1.09	.52	.24	.27	76.1	70.6	S10
6	15	48	.39	.33	.90	-.78	.84	-.87	.38	.28	73.9	68.8	S6
12	16	48	.28	.32	1.02	.18	1.06	.42	.26	.29	65.2	67.0	S12
13	18	48	.08	.32	1.01	.14	1.13	.91	.27	.30	69.6	63.7	S13
19	18	48	.08	.32	1.04	.49	1.16	1.12	.24	.30	60.9	63.7	S19
8	20	48	-.12	.31	1.13	1.49	1.13	1.09	.20	.32	65.2	61.8	S8
2	21	48	-.21	.31	.94	-.76	1.00	.05	.37	.33	69.6	61.7	S2
9	21	48	-.21	.31	.84	-2.01	.81	-1.72	.48	.33	78.3	61.7	S9
7	22	48	-.31	.31	.95	-.63	.91	-.78	.39	.34	63.0	61.6	S7
11	22	48	-.31	.31	1.14	1.63	1.13	1.12	.22	.34	45.7	61.6	S11
14	24	48	-.50	.31	.96	-.41	.95	-.43	.39	.35	65.2	62.4	S14
4	25	48	-.59	.31	.99	-.13	.95	-.41	.38	.36	65.2	62.9	S4
18	25	48	-.59	.31	1.06	.74	1.12	1.04	.29	.36	60.9	62.9	S18
5	27	48	-.79	.31	.83	-1.76	.80	-1.76	.53	.37	73.9	64.8	S5
3	31	48	-1.20	.33	.98	-.09	.96	-.22	.42	.40	71.7	70.4	S3
MEAN	18.2	48.0	.24	.40	1.00	.0	1.01	.0			68.5	66.9	
P.SD	7.1	.0	1.20	.32	.09	.9	.12	.9			8.3	6.1	

Figure 2. Item Measure

Figure 2. above reveals the level of student ability obtained through the analysis of the person measure table, which provides detailed logit information for each individual or student. Further analysis is used to evaluate the fit order of each question item so that it can assess the suitability and feasibility of the question. Based on the indicator 0.5 < MNSQ < 1.5, it is known that from question number 1 to number 20, there are no questions classified as misfit order. The indicator -2.0 < ZSTD < 2.0 shows that from question number 1 to question number 20, no questions fall into the misfit order category. Meanwhile, based on the indicator 0.4 < Pt Mean Corr < 0.85, it is known that apart from questions 5, 3, and 9 tend to be less suitable (Widhiarso, 2015). However, because the limitations of the ZSTD indicator have been met, these questions can still be considered suitable for use. Thus, there are no questions that need to be replaced or changed. Based on the measure order analysis, the level of difficulty of the questions can be categorized as follows: (a) the most difficult questions are questions number 16, 20, 17; (b) the questions that are considered difficult are questions number 15, 1, 10, 6, 12, 13, 19, and 8; (c) the questions that are considered easy are questions number 2, 9, 7, 11, 4, 18 and 5; and (d) the most straightforward question is question number 3.

The results of the person measure table analysis provide an overview of students' individual abilities through logit values that reflect the level of ability of each. Evaluation of the fit order of the test items is carried out to assess the feasibility and suitability of the test instrument. Based on the Mean Square Outfit (MNSQ) indicator with the criteria of 0.5 < MNSQ < 1.5, all questions from number 1 to number 20 show appropriate results so that no questions are included in the misfit category. In addition, the analysis results using the indicator -2.0 < ZSTD < 2.0 also show that all questions are within the feasibility range. This shows that the questions tested have consistent characteristics in measuring students' abilities according to the RASCH model, as explained by Widhiarso (2015).

However, based on the indicator 0.4 < Pt Mean Corr < 0.85, there are several questions, such as questions 5, 3, and 9, which are considered less appropriate or unfit. However, because the ZSTD values of these questions still meet the eligibility criteria, these questions are still considered usable in the test. This finding indicates that the test instrument is reliable and does not require changes to the composition of the questions. This is based on the RASCH modeling principle, which prioritizes testing the suitability between empirical data and theoretical models, thereby minimizing the potential for bias in measurement (Fan & Bond, 2019).

The difficulty level of the questions can be categorized based on the results of the measure order analysis. The questions with the highest difficulty level are questions 16, 20, and 17, which can be used as a reference to identify students with high abilities. Questions classified as difficult include questions 15, 1, 10, 6, 12, 13, 19, and 8, while questions with a low level of difficulty include questions 2, 9, 7, 11, 4, 18, and 5. The most straightforward question is question 3. This classification is essential to evaluate the distribution of the difficulty level of questions in supporting a more effective learning process. With an even distribution of difficulty levels, the test can provide more comprehensive information about students' abilities at various levels and help develop learning strategies that suit their needs.

3.3 Scalogram Analysis

```
GUTTMAN SCALOGRAM OF RESPONSES:
Person |Item
       |   11 1   111  11121
       |35484712983926105706
       |--------------------
    37 +11111101101111010100  37IC
    26 +11101111001101011100  26IC
    36 +11101101111001010100  36CC
     3 +11111101010110100000  03SA
     6 +11111001110000101000  06NA
    12 +00010101111111000010  12AA
    15 +11111101010110000000  15MA
    32 +11101001111000010100  32MC
    33 +11101100011001101000  33MC
    40 +01011111101110000000  40MC
    44 +11010111101011000000  44RC
    45 +11001001110011001010  45AC
    13 +10110100010001101100  13FA
    16 +10110010100000111100  16AA
    20 +11110000110101100000  20AA
    23 +11110010100000111000  23HA
    31 +01011010100110100010  31RC
    38 +10011100101000011010  38NC
    39 +01011011101110000000  39JC
    42 +01011111100110000000  42BC
    43 +11000111101011000000  43AC
    14 +00100110001101101000  14ZA
    19 +10110000100000111010  19DA
    21 +11101011100100000000  21MA
    22 +10110010100000111000  22AA
    24 +11101111000001000000  24DA
    30 +11011010010100100000  30KC
```

Figure 3. Scalogram

In RASCH modeling, students' response patterns can be analyzed based on their ability levels, from the highest to the lowest. This analysis refers to the ordering of questions based on their difficulty levels, where the questions are arranged horizontally, with the most straightforward questions on the far left. Based on further analysis using a scalogram, students can be grouped into three categories: careless, lucky guess, and cheating. The careless category includes students who can answer difficult questions but need help to solve easier questions. Examples are student 12AA and student 9DA.

Meanwhile, the lucky guess category refers to students who cannot solve the most straightforward questions but succeed in answering difficult questions, such as student 31RC. Based on the definition in RASCH modeling, students with low ability should not have the opportunity to answer difficult questions. Therefore, the correct answers from students on difficult questions are thought to result from guessing. The last category, cheating, refers to students with identical answer patterns, such as students 20AA and 23HA. RASCH modeling shows that students with different ability levels should not have the same opportunity to answer questions with identical correct patterns. Therefore, the similarity in answer patterns between students 20AA and 23HA indicates cheating activities (Yulianto & Widodo, 2020; Suryani, 2018).

RASCH modeling systematically analyzes students' response patterns based on their ability levels, from highest to lowest. This pattern is integrated with ordering questions by difficulty level, with the most straightforward questions placed on the leftmost position in the scalogram. Based on this analysis, students can be grouped into three main categories: careless, lucky guessers, and cheaters. These findings reflect the importance of evaluating response patterns to improve test validity, as inconsistent patterns may indicate behavior inconsistent with the measurement model.

The careless category includes students who can answer complex questions but fail on easy questions, as seen in students 12AA and 9DA. This can be caused by carelessness, lack of attention, or fatigue during the test. According to Bond and Fox (2013), this kind of behavior can reduce the reliability of the test, so it requires a more contextual learning approach to improve students' consistency in answering questions. In contrast, the lucky guess category refers to students who cannot solve easy questions but succeed on difficult questions, as seen in students 31RC. RASCH modeling suggests that students with low ability should not have the opportunity to answer difficult questions. Therefore, the correct answers to difficult questions in this case are suspected to be the result of guessing or luck, which reflects the importance of improving the design of questions to minimize the possibility of answers that do not reflect actual ability.

The last category, mutual cheating, is characterized by the similarity of answer patterns between two or more students, as seen in students 20AA and 23HA. RASCH modeling indicates that students with different abilities should have different response patterns. Significant similarities in answer patterns indicate the possibility of cheating activities during the test. This finding is in line with the research of Yulianto and Widodo (2020), which states that cheating can reduce test results' integrity. To overcome this, it is essential to implement preventive measures, such as strict supervision during the test or using more varied question formats to minimize opportunities for cheating. These efforts can increase the validity and reliability of test results as a learning evaluation tool.

The results of the Wright Map analysis indicate a significant variation in students' abilities in understanding basic mathematical concepts. While the majority of students performed well on easier questions, their performance on more difficult questions remained suboptimal. This finding aligns with Widhiarso (2015), who emphasized the necessity of diverse instructional strategies to address differences in students' cognitive abilities. However, unlike previous studies that focused solely on traditional learning models (Syadiah & Hamdu, 2020), this study provides evidence that the flipped classroom approach may not fully bridge the gap between students with different ability levels.

A critical observation from the Wright Map distribution is that students with higher ability levels demonstrated consistent mastery of complex mathematical concepts, whereas lower-achieving students struggled even with fundamental topics. This pattern raises concerns about the effectiveness of flipped classrooms for all student types. Research by Lage et al. (2000) suggested that flipped learning is particularly beneficial for self-motivated learners, but its effectiveness may be limited for students who require more structured guidance. Therefore, additional interventions, such as blended learning or differentiated instruction, should be considered to ensure equitable learning outcomes.

Furthermore, external factors may influence students' performance. Factors such as prior knowledge, digital literacy, and home learning environments can significantly impact the success of flipped classrooms (Bishop & Verleger, 2013). The present study does not account for these variables, which could be an important area for future research. Additionally, the results indicate that certain students answered difficult questions correctly while struggling with easier ones, suggesting potential issues such as test-taking anxiety, guessing, or instructional gaps. Addressing these challenges requires a more adaptive teaching approach that accommodates individual learning needs.

In terms of pedagogical implications, these findings highlight the need for elementary school teachers and PGSD lecturers to implement more targeted instructional strategies. While the flipped classroom approach can facilitate active learning, its application in teaching number concepts should be complemented with formative assessments and scaffolding techniques to support struggling students. Moreover, professional development programs should be designed to equip educators with strategies for integrating flipped learning effectively in diverse classroom settings. By addressing these issues, flipped classrooms can become more inclusive and supportive for students with varying abilities.

4. CONCLUSION

This study found that the implementation of the flipped classroom model, when analyzed through the RASCH model, effectively enhanced students' understanding of number concepts while revealing diverse response patterns among learners. The RASCH-based scalogram analysis identified three distinct response behaviors: *careless* (students answering difficult items correctly but failing on easier ones), *lucky guessers* (students who answered difficult items correctly despite low overall ability), and *cheating* (students with identical response patterns despite differing abilities). These findings suggest that while the flipped classroom can support conceptual learning, variations in student behavior and ability persist. However, this research is limited by its small sample size (39 students from a single class), lack of a control group, and focus on a single mathematical topic, which may limit the generalizability of the findings. Future research should involve larger and more diverse samples, include control or comparison groups, and explore how external factors—such as digital literacy, motivation, or learning environment—interact with flipped learning models to affect outcomes. Additionally, combining RASCH analysis with qualitative methods may offer deeper insights into the causes behind response anomalies and support the development of more adaptive and equitable instructional strategies.

REFERENCES

Aledya, V. (2020). *Kemampuan pemahaman konsep matematika pada siswa*. Universitas Negeri Medan.

Asri Fauzi, A., Rahmatih, A. N., & Lestari, F. H. (2022). Analisis efektivitas model pembelajaran blended learning ditinjau dari hasil belajar geometri mahasiswa guru sekolah dasar. *Elementary Education*, 5(1), 43–52.

Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.

Dwi Putra, H., Thahiram, N. F., Ganiati, M., & Nuryana, D. (2018). Kemampuan pemecahan masalah matematis siswa SMP pada materi bangun ruang: Development of project-based blended learning model to support student creativity in designing mathematics learning in elementary school. *JIPM (Jurnal Ilmiah Pendidikan Matematika)*, 6(2), 82–90. http://journal.unipma.ac.id/index.php/jipm

Dwinata, A. (2019). Analisis kemampuan pemecahan masalah menggunakan pemodelan Rasch pada materi permutasi dan kombinasi. *PRISMA, Prosiding Seminar Nasional Matematika*, 2, 124–131. https://journal.unnes.ac.id/sju/index.php/prisma/article/view/28899

Fan, J., & Bond, T. (2019). Applying Rasch measurement in language assessment: Unidimensionality and local independence. In V. Aryadoust & J. Fox (Eds.), *Quantitative data analysis for language assessment: Volume I* (pp. 83–102). Routledge.

Fauzi, A., Sawitri, D., & Syahrir, S. (2020). Kesulitan guru pada pembelajaran matematika di sekolah dasar. *Jurnal Ilmiah Mandala Education*, 6(1), 142–148.

Rahayu, E. (2021). Problema kesulitan siswa sekolah dasar dalam pembelajaran geometri. *At-Tàlim: Jurnal Pendidikan*, 7(1), Article 2548–4419.

Rahayu, L. D., & Kusuma, A. B. (2019). Peran pendidikan matematika di era globalisasi. *Prosiding Sendika*, 5(1), 534–541. http://eproceedings.umpwr.ac.id/index.php/sendika/article/view/801

Rahmat, A. A., Hamdu, G., Nur'aeni, E., & Lidinillah, D. A. M. (2020). Pengembangan soal tes tertulis berbasis STEM dengan pemodelan Rasch di sekolah dasar. *Jurnal Didika: Wahana Ilmiah Pendidikan Dasar*, 6(1), 29–40. https://doi.org/10.29408/didika.v6i1.2197

Rusilowati, A. (2018). Asesmen literasi sains: Analisis karakteristik instrumen dan kemampuan siswa menggunakan teori tes modern Rasch model. *Prosiding Seminar Nasional Fisika Universitas Riau Ke-3*, 2–15. https://snf.fmipa.unri.ac.id/wp-content/uploads/2019/03/0.-300B-2-15NI.pdf

Samintono, B., & Widhiarso, W. (2013). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial*. Trim Komunikata Publishing House.

Sariningsih, R., & Purwasih, R. (2017). Pembelajaran problem based learning untuk meningkatkan kemampuan pemecahan masalah matematis dan self-efficacy mahasiswa calon guru. *JNPM (Jurnal Nasional Pendidikan Matematika)*, 1(1), 163. https://doi.org/10.33603/jnpm.v1i1.275

Sugiyono. (2013). *Metode penelitian pendidikan: Pendekatan kuantitatif, kualitatif dan R&D*. Alfabeta.

Suryani, Y. E. (2018). Aplikasi Rasch model dalam mengevaluasi Intelligenz Structure Test (IST). *Psikohumaniora: Jurnal Penelitian Psikologi*, 3(1), 73–86. https://doi.org/10.21580/pjpp.v3i1.2052

Syadiah, A. N., & Hamdu, G. (2020). Analisis Rasch untuk soal tes berpikir kritis pada pembelajaran STEM di sekolah dasar. *Premiere Educandum: Jurnal Pendidikan Dasar dan Pembelajaran*, 10(2), 138–148. https://doi.org/10.25273/pe.v10i2.6524

Tarigan, E. F., Nilmarito, S., Islamiyah, K., Darmana, A., & Suyanti, R. D. (2022). Analisis instrumen tes menggunakan Rasch model dan software SPSS 22.0. *Jurnal Inovasi Pendidikan Kimia*, 16(2), 92–96. https://doi.org/10.15294/jipk.v16i2.30530

Tyas, E. H., Hamdu, G., & Pranata, O. H. (2020). Analisis soal pilihan ganda dengan menggunakan pemodelan Rasch untuk mengukur kemampuan siswa dalam mengurutkan bilangan pecahan di sekolah dasar. *Pedadidaktika: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, 7(2), 1–12.

Widhiarso, W., & Kozeny, J. (2013). Applying confirmatory factor analysis to evaluate the fit of measurement models in education and psychological assessment. *Journal of Educational Sciences and Psychology*, 3(2), 1–12.

Yulianto, A., & Widodo, A. (2020). Disclosure of difficulty distribution of HOTS-based test questions through Rasch modeling. *Indonesian Journal of Primary Education*, 4(2), 197–203. https://doi.org/10.17509/ijpe.v4i2.29318

Yuwono, T., Supanggih, M., & Ferdiani, R. D. (2018). Analisis kemampuan pemecahan masalah matematika dalam menyelesaikan soal cerita berdasarkan prosedur Polya. *Jurnal Tadris Matematika*, 1(2), 137–144. https://doi.org/10.21274/jtm.2018.1.2.137-144